

STIC-Biotech/ChemLib

93116

From: Rao, Manjunath N.
Sent: Monday, May 05, 2003 8:52 AM
To: STIC-Biotech/ChemLib
Subject: Sequence search request for 09/868,328

From: Manjunath N. Rao
Art Unit 1652, Room 10A11
Mail Box in Room 10D 01
Phone: 306-5681

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

RECEIVED
MAY -5 2003
STIC/CHEN, D. (STIC)

Date: 5-5-03

Please search the following as soon as possible for application with serial number **09/868328**

1. SEQ ID NO: 2 and 3 against all commercial nucleic acid databases including issued patents database and pending application database and provide a print of all results.
2. SEQ ID NO:1, 4 against all commercial protein databases including issued patents database and pending application database and provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.
Biotechnology Patent Examiner
Art Unit 1652, Room 10A11
Mail Box in 10D01
Crystal Mall 1, USPTO.

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 5/6
Date Completed: 5/12
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:23:32 ; Search time 35.0673 Seconds
(without alignments)
1979.723 Million cell updates/sec

Title: US-09-868-328B-1

Perfect score: 2845

Sequence: 1 MTPAISRAVLQAGAGALA.....TDGPAHFTGIVVREIQAI 521

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SID22/qcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID22/qcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID22/qcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID22/qcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID22/qcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID22/qcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID22/qcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID22/qcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID22/qcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID22/qcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID22/qcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID22/qcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID22/qcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID22/qcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID22/qcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID22/qcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID22/qcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID22/qcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID22/qcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID22/qcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID22/qcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID22/qcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID22/qcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2845	100.0	521	22 AAB82301	Arthrobacter ureaf
2	2355	82.8	517	20 AAY04105	Arthrobacter nicot
3	2355	82.8	517	23 ABB10074	Levan fructotransf
4	2268	79.7	484	20 AAY04104	Arthrobacter nicot
5	2268	79.7	485	23 ABB10073	Levan fructotransf
6	499.5	17.6	943	21 AAY91929	Paenibacillus mace
7	468.5	16.5	1277	21 AAY91928	Paenibacillus pabu
8	461	16.2	923	21 AAY91927	Paenibacillus amyl
9	427	15.0	516	21 AAU76758	Novel recombinant
10	416.5	14.6	533	19 AAU44863	C. utilis INV1 pro

11	389.5	13.7	490	16 AAR72566	Penicillium purpur
12	386	13.6	750	20 AAY25346	Bacillus sp. L7 en
13	386	13.6	750	20 AAY24924	Bacillus L7 endo-1
14	366	12.9	556	15 AAR56256	Inulinase preprote
15	359.5	12.6	532	5 AAP40020	Saccharomyces cere
16	359.5	12.6	532	12 AAR12145	Preinvertase. Sac
17	348	12.2	581	20 AAY17500	Schizosaccharomyce
18	348	12.2	581	20 AAY08399	S. pombe invertase
19	348	12.2	581	21 AAB03200	Schizosaccharomyce
20	292.5	10.3	671	23 AAU74396	Fructosyl transfer
21	277.5	9.8	636	16 AAR75187	Tomato plant inver
22	277.5	9.8	636	21 AAB21098	Wild-type tomato 1
23	274.5	9.6	636	13 AAR27007	Tomato vacuolar in
24	274.5	9.6	636	14 AAR35102	Tomato acid invert
25	266	9.3	583	22 AAE04416	Lycopersicon penne
26	266	9.3	648	23 AAU74395	Fructosyl transfer
27	261.5	9.2	1116	20 AAY43179	S. rochei strain E
28	258.5	9.1	645	23 ABB77844	Amino acid sequenc
29	244.5	8.6	479	23 ABB27213	Streptococcus poly
30	243	8.5	670	21 AAY85665	Corn invertase pro
31	241	8.5	553	15 AAR53922	Tomato acid invert
32	241	8.5	553	16 AAR82848	Tomato acid invert
33	239.5	8.4	1487	16 AAR74634	Bacillus circulans
34	235.5	8.3	492	23 ABB39050	Staphylococcus epi
35	235.5	8.3	591	21 AAG21596	Arabidopsis thalia
36	231.5	8.1	587	23 ABB92634	Herbicidally activ
37	223.5	7.9	569	23 ABB92300	Herbicidally activ
38	222.5	7.8	554	23 AAU74399	Fructosyl transfer
39	222	7.8	591	23 ABB91362	Herbicidally activ
40	220	7.7	581	23 ABB92301	Herbicidally activ
41	220	7.7	588	21 AAG50247	Arabidopsis thalia
42	205.5	7.2	630	17 AAU00455	1-Sucrose:sucrose
43	205.5	7.2	637	19 AAU77034	Artichoke sucrose
44	202.5	7.1	429	22 AAB83190	Corynebacterium th
45	202.5	7.1	433	22 AAG92651	C glutamicum prote

ALIGNMENTS

RESULT 1
AAB82301
ID AAB82301 standard; Protein: 521 AA.
XX AAB82301;
AC AC
XX 09-JUL-2001 (first entry)
DT Arthrobacter ureafaciens levan fructotransferase.
DE Levan fructotransferase; difructose dianhydride IV; sweetenur.
KW Arthrobacter ureafaciens.
XX Arthrobacter ureafaciens.
OS
FH Key Location/Qualifiers
FT Peptide 1..33
FT /label= Signal_peptide
FT Protein 34..521
FT /label= Mature_protein
XX
PN WO200129185-A1.
XX
PD 26-APR-2001.
XX
PE 19-OCT-2000; 2000WO-KR01183.
XX
PR 19-OCT-1999; 99KR-0045302.
XX
PA (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
PA (REAL-) REALBIOTECH LTD.
XX
PI Rhee S, Song K, Kim C, Ryu E, Lee Y;
XX

DR WPI; 2001-308483/32.
 XX N-PSDB; AAF30918, AAF30919.
 PT Producing difructose dianhydride IV from sucrose, involves reacting
 PT sugar solution in the presence of levansucrase to produce levan, and
 PT reacting levan solution in the presence of levan fructotransferase to
 PT produce DFA IV -
 XX
 PS Claim 2; Page 47; 72pp; English.
 XX
 CC The present sequence is that of *Arthrobacter ureafaciens* K2032
 CC levan fructotransferase (see AAB82301). The enzyme can be
 CC obtained by cultivation of claimed *Escherichia coli* JUD81
 CC (KCTC 0877BP), which carries claimed expression vector pUDFA81
 CC comprising the levan fructotransferase gene (see AAF30918). A
 CC claimed process for producing difructose dianhydride IV from
 CC sucrose comprises subjecting sugar solution to reaction at room
 CC temperature or lower in acidic buffer of pH 3.0-7.0 in the presence
 CC of a levansucrase derived from *Zymomonas mobilis* to produce levan,
 CC purifying the levan from the reaction solution, and subjecting it
 CC to reaction at 25-30 degree C for 3-10 hours in acidic buffer of pH
 CC 3.0-7.0 in the presence of levan fructotransferase, preferably
 CC obtained from *E. coli* JUD81. The product is useful as a low-calorie
 CC sweetener.
 XX
 PS Sequence 521 AA;

Query Match 100.0%; Score 2845; DB 22; Length 521;
 Best Local Similarity 100.0%; Pred. No. 7.1e-241;
 Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTPAISRAVLQAGAGALALIFGGVPPAARASAPGSLRAVYHMTPPSGWLCDDPQRPVT 60
 DB 1 MTPAISRAVLQAGAGALALIFGGVPPAARASAPGSLRAVYHMTPPSGWLCDDPQRPVT 60
 QY 61 THGAYQLYLHSDQNGPGGWDHASTTGDVAFTHHGTGTMPLRDPFVWSSAVVGTANTA 120
 DB 61 THGAYQLYLHSDQNGPGGWDHASTTGDVAFTHHGTGTMPLRDPFVWSSAVVGTANTA 120
 QY 121 GFGAGAVVALATOPTDGVKRYQYLYWSTDDGFTTALPDPIVNTDGRAATPAEIE 180
 DB 121 GFGAGAVVALATOPTDGVKRYQYLYWSTDDGFTTALPDPIVNTDGRAATPAEIE 180
 QY 181 AEWFRDPKTHWDTPARGEWCVIGRLRYAAYFTSPNLRDWTLRNFDYPNHALGGIECPDL 240
 DB 181 AEWFRDPKTHWDTPARGEWCVIGRLRYAAYFTSPNLRDWTLRNFDYPNHALGGIECPDL 240
 QY 241 FEITADDGTRHHWVLAASMDAYGIGLPMYAYWTGTWDGQFHADDLTPOWLDGWWDYAA 300
 DB 241 FEITADDGTRHHWVLAASMDAYGIGLPMYAYWTGTWDGQFHADDLTPOWLDGWWDYAA 300
 QY 301 VTWPSIDAPETKRLATAMNNWKYAAARDVPTDASDGYNGQNSIVRELRLARQPGWYTLL 360
 DB 301 VTWPSIDAPETKRLATAMNNWKYAAARDVPTDASDGYNGQNSIVRELRLARQPGWYTLL 360
 QY 361 STPVAALTNYVTATTTLPDRTVDSAVLPWNGRAYEIELDIAWDTATNVGISVGRSPDGT 420
 DB 361 STPVAALTNYVTATTTLPDRTVDSAVLPWNGRAYEIELDIAWDTATNVGISVGRSPDGT 420
 QY 421 RHTNICKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILDVDTQSVEVFNA 480
 DB 421 RHTNICKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILDVDTQSVEVFNA 480
 QY 481 GHTVLSQQVHFAEGDTGISLYTDGGPAHFTGIVVREIGQAI 521
 DB 481 GHTVLSQQVHFAEGDTGISLYTDGGPAHFTGIVVREIGQAI 521

RESULT 2
 AAY04105
 ID AAY04105 standard; Protein; 517 AA.
 XX
 AC AAY04105;

XX 10-JUN-1999 (first entry)
 DT
 XX
 DE Arthrobacter nicotinovorans levan fructotransferase protein #2.
 KW Arthrobacter nicotinovorans; levan fructotransferase.
 XX
 OS Arthrobacter nicotinovorans.
 PN JP11069978-A.
 XX
 PD 16-MAR-1999.
 XX
 PF 28-AUG-1997; 97JP-0232421.
 PR 28-AUG-1997; 97JP-0232421.
 XX
 PA (NIOC) NIPPON OIL CO LTD.
 XX
 DR WPI; 1999-247463/21.
 DR N-PSDB; AAX19827.
 XX
 PT Levan fructotransferase gene - for recombinant production of levan
 PT fructotransferase
 XX
 PS Claim 3; Page 8-9; 14pp; Japanese.
 CC
 CC The present sequence represents *Arthrobacter nicotinovorans* levan
 CC fructotransferase. The present invention also describes a method
 CC for the preparation of levan fructotransferase in which a transformant
 CC is cultured in a medium and levan fructotransferase is collected from
 CC the culture. The method can prepare levan fructotransferase in a
 CC large amount.
 XX
 PS Sequence 517 AA;
 QY
 DB 1 MTPAISRAVLQAGAGALALIFGGVPPAARASAPGSLRAVYHMTPPSGWLCDDPQRPVT 60
 DB 1 MTYDISRRALOGAGALALFMSNAIPVAHAQA--SLRAIYHMTPPSGWLCDDPQRPVH 58
 QY 61 THGAYQLYLHSDQNGPGGWDHASTTGDVAFTHHGTGTMPLRDPFVWSSAVVGTANTA 120
 DB 59 TNGAYQLYLHSDQNGPGGWDHATTGCGVSYTHHGVWMPQDPFVWSSAVVDTANTA 118
 QY 121 GFGAGAVVALATOPTDGVKRYQYLYWSTDDGFTTALPDPIVNTDGRAATPAEIE 180
 DB 119 GFGAGAVIALATQPTDG--KFQEQYLYWSTDDGYSFTALPDPIVNTDGRATTAEVEN 176
 QY 181 AEWFRDPKTHWDTPARGEWCVIGRLRYAAYFTSPNLRDWTLRNFDYPNHALGGIECPDL 240
 DB 177 AEWFRDPKTHWDTPARGEWCVIGRLRYAAYFTSPNLRDWTLRNFDYPNHALGGIECPDL 236
 QY 241 FEITADDGTRHHWVLAASMDAYGIGLPMYAYWTGTWDGQFHADDLTPOWLDGWWDYAA 300
 DB 237 FEMTAGDGTTHWVFGASMDAYSIGLPMYAYWTGWSNGTAFADNLTPOWLDGWWDYAA 296
 QY 301 VTWPSIDAPETKRLATAMNNWKYAAARDVPTDASDGYNGQNSIVRELRLARQPGWYTLL 360
 DB 297 VTWPAVEAPETKRLATAMNNWKYAAARNVPTDASDGYNGQNSITRELRLERQSGGWYTLL 356
 QY 361 STPVAALTNYVTATTTLPDRTVDSAVLPWNGRAYEIELDIAWDTATNVGISVGRSPDGT 420
 DB 357 STPVPALSNYATSSITLPDRTVNGSEVLWPSGRAYEIELDISWDTAANVGVSGRSSDGS 416
 QY 421 RHTNICKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILDVDTQSVEVFNA 480
 DB 417 RHTNICKYGDELYVDRASSEQSGYALAPYTRAAAPIDANARSVHLRIFVDQTSVEVFNS 476
 QY 481 GHTVLSQQVHFAEGDTGISLYTDGGPAHFTGIVVREIGQAI 521

Db 776 QKTT--VAYRVDTQNMVFDRTTSGVFSFSDLFTKVKHEASLKPENQKVKLRIFVDESSVEV 833
QY 477 FVNGHTVLSQOVHFAEGDGTGSLYTDGGPAHFTGIVV 514
Db 834 FGNGKVVFSVDVIFPDAGRAMAFYSLGGEVKVSSMKV 871

RESULT 8
AAV91927
ID AAV91927 standard; Protein; 923 AA.
AC AAV91927;
XX 19-JUL-2000 (first entry)
XX Paenibacillus amylolyticus 2,6-beta-D-fructan hydrolase.
XX 2,6-beta-D-fructan hydrolase; anticaric; neuroprotective; biofilm;
KW hydrolyze slime; detergent; fructose production; sweetener.
XX Paenibacillus amylolyticus.
XX Key Location/Qualifiers
FT Protein 32..923
FT /label= mature_protein
XX WO200017331-A1.
XX 30-MAR-2000.
XX 17-SEP-1999; 99WO-DK00495.
XX 18-SEP-1998; 98DK-0001173.
XX 09-DEC-1998; 98DK-0001623.
XX (NOVO) NOVO-NORDISK AS.
XX Moller S, Johansen C, Schaefer T, Ostergaard PR, Hoeck LH;
XX WPI: 2000-283564/24.
XX N-PSDB; AAA08533.
XX New polypeptides with 2,6-beta-D-fructan hydrolase activity, useful
PT e.g. for degrading microbial biofilm, hydrolyzing slime, production of
PT fructose oligomers and treatment of dental plaque
XX Claim 1; Page 107-110; 123pp; English.
XX AAV91927-29 are polypeptides from Paenibacillus amylolyticus, P. pabuli
CC and P. macerans, respectively, which have 2,6-beta-D-fructan hydrolase
CC activity. These are preferred enzymes of the invention. The enzymes
CC can be used to remove microbial biofilm from surfaces, particularly for
CC reduction of biocorrosion, for treatment of dental plaque and for
CC treatment of multiple sclerosis (eliminating biofilm from the lungs; to
CC hydrolyze slime; in detergents and for production of fructose oligomers
CC from levan or phlelin (useful as sweeteners or for conversion to
CC fructose). The nucleic acids that encode the enzymes are useful for
CC recombinant production and as source of probes and primers for
CC identification and cloning of related fructan hydrolases (all claimed).
XX Sequence 923 AA;
Query Match 16.2%; Score 461; DB 21; Length 923;
Best Local Similarity 29.5%; Pred. No. 2.8e-31;
Matches 150; Conservative 74; Mismatches 223; Indels 62; Gaps 23;

QY 40 RAYVHTPPSGWCLDQRPVTHGAYQLYVLSQDQNGPGWDHASTDGVAFTHGVTVM 99
Db 392 RPQYHTPIRGASDPNGLYFEGEYHLF--HQD----GGTWAHVAVSKDMLNKKRLPIAL 445
QY 100 PLRPDPFVWSGSVAVGTANTAGF-----GAGAVVALATQPTDGVKRYOEQLYWSIDGGF 154

Db 446 PWNHGHVWSGSVAVADMTNAGSLFGDGGKGLIAYTTFNPDSPNGNORIGLAYSKDQGR 505
QY 155 TETALPD--PVIVNTDGRAATTPAEIEAEN--FRDPKTHWDTARGEWVCVIGRLRYAAFYT 212
Db 506 TWEYSKERPIVENPGKSGN-----EAGNDRDPKVIDDENNRWVWVSGGDHIREYT 560
QY 213 SPNLRDWTLRRNFDYPNHALGGI--ECPDLFEITADGCT--RHVYLAASMDAYGIGLPMY- 268
Db 561 STNLLDWTLLDNMGYGDYVRGGVWECPDLPV-DGTSQKWKVMMISTGAN-----PKTG 615
QY 269 ---YAYWTG--TWDEGFHADDLTPOWL--DWGWDWAAVWPSIDAPETKRLAIAMNN 321
Db 616 GSDAEYFIGHLTAG-KFVNDNPAGKVLRTDFGKEFYASMFANM--PDHRTVMAMWNTN 672
QY 322 WKYAADVPDPTDASDYGNGQNSIVRELRLARQPGGWYLLSTFPVAAALNYVTATTTLPDRT 381
Db 673 WDYPFA-FPT---SNWKGELTIPREVSIVTTEDEG-IRMVQSPIKEL-----ESLRKPLYS 722
QY 382 VDSAVLPWNGR-----AYEIELDI---AMDATNVGISVGRSPDGTFRHNIKGYKA 430
Db 723 ASNKSVPSSGNLLKGIISGAYEIEAIEIPETSTVTEFGFNIREG--ANQKTWVGKAS 780
QY 431 D--LYVDRGPDLAGYSLAPYSRAAAPIDFGARSVHLRLIVDTQSVEFVNAGHTVLSQ 488
Db 781 DSRMVDRTASGETDFSNLFSKKEAPTQMNENRKRILVDESSVEAFGNDGKVVFSV 840
QY 489 VFHAEGDTGISLYTDGGPAHFTGIVVREI 517
Db 841 IFPDPASRAMSFYVKGNGVNVVSLKVHQL 869

RESULT 9
AAU76758
ID AAU76758 standard; Protein; 516 AA.
XX AAU76758;
AC AAU76758;
DT 20-MAY-2002 (first entry)
XX Novel recombinant Aspergillus endoinulinase associated protein.
XX Recombinant expression vector; Aspergillus endoinulinase;
KW transformed microorganism.
XX Unidentified.
XX KR2000004834-A.
XX 25-JAN-2000.
XX 10-JUL-1998; 98KR-0027920.
XX 10-JUN-1998; 98KR-0021432.
XX (UHMT/) UHM T B.
PA (CHAE/) CHAE G S.
PA (HOIT/) HOIT J V.
XX Uhm TB, Chae GS;
XX WPI: 2000-677409/66.
DR N-PSDB; ABK10615.
XX Novel recombinant Aspergillus endoinulinase gene derived expression
PT vector and transformed microorganism - NoAbstract
XX Claim 2; Page 10; 12pp; Korean.
XX The present invention relates to a new recombinant Aspergillus
CC endoinulinase gene derived expression vector and also describes a
CC transformed microorganism. The present amino acid sequence represents
CC the novel recombinant Aspergillus endoinulinase gene derived expression
CC vector protein of the invention.

```
XX SQ Sequence 516 AA;
Query Match 15.0%; Score 427; DB 21; Length 516;
Best Local Similarity 27.2%; Pred. No. 1.2e-28;
Matches 145; Conservative 92; Mismatches 221; Indels 76; Gaps 21;
Qy 24 GGAVPPAARASAGSLRAYVHTPPSGWLCDPQRPVTHGAYQLYLHSDONNGPGG--W 81
Db 16 GLTLPSQAQSN---DYRSPSYHTPDQYWNNEPGLIKIGTSTWHLFFQHNPNTANWGNICW 72
Qy 82 DHASTDGVATHTHCTVPLRDPFVWSGSVAVGTANTAGFAGAGA---VVALATOPTDGV 138
Db 73 GHATSTDLHMAHKPTALADENGVEAFTGTAYDPNNTSGLGDSANPPYLAFTGYTTS- 131
Qy 139 RYQEOYLWSTGDGFTTALPDVIVNTDGRAATTAEIENAEWFRDPKIHWDTFARGEW 198
Db 132 SOTQORLAFSVNGATWTKFQGNPIIISTQEA---PHDITGGLESRDPKVFHFRQSGNW 188
Qy 199 VCVI--GRLRYAAFTSPNLRDWTLLRRNFDYPNHALGGI-----ECPDLFEITADDT 249
Db 189 IMVLAHGGODKLSFWTSADTINWTWQS--DLKSTSLNGLSDDITGWEPDMFELPV-EGT 245
Qy 250 RH--WVL-----AASMDAIGLPMYIAYTWTGDEQFHAD--DLTPQWLQNGWDWYAAV 301
Db 246 BETTWVNMTPAEGSPAGGNGVLAI---TGSFDCGKSFADPVDASTMWLQNGRDFDGL 301
Qy 302 TPWSDIDAPETRLATAWNNKYYAARDVPTDASDGYNGQNSIVRELRLAROPGGWYTLIS 361
Db 302 SWNVPSADGRRIIAVWNSY---GSPNPTTT---WKMLGFPPTLSL-KKVGTOOHEVQ 354
Qy 362 TPVAALTYVATTTLPTORTYDGSVAVLPWNGRAYEIELDIAW--DTATNVGISVGRSPDG 419
Db 355 QPITELDTFISTSLQILANQITTPGQTLSSIRGTALDVRVAFYDPDAGSVLSIAVRK--G 411
Qy 420 TRHNTIGKY---GADLYVDRGPSDLAGYSLAPYSRAAPIIDPGARSVH-----464
Db 412 ASEQTVIKYQSDATLSVDRTESGDISY-----DPAAGGVHTAKLEEDGTGLV 459
Qy 465 -LRILVDPOSVEVFVNAAGTIVLSQCVHFAEGDGTGISLTDGPGPAHFTGIVVREI 517
Db 460 SIRVLVDTCSEVVEVGQGEANVISDLIFPSSDSDGLALEVTGNAVLSQVDVRSV 513
RESULT 10
AAW44863
ID AAW44863 standard; Protein; 533 AA.
XX AC AAW44863;
XX DT 29-SEP-1998 (first entry)
XX DE C. utilis INV1 protein.
XX KW URA3; HIS3; INV1; orotidine 5'-monophosphate decarboxylase; enzyme;
XX KW biosynthesis; ribonucleotide; uracil; histidine; auxotroph; invertase;
XX KW imidazole-glycerol-phosphate dehydratase; beta-fructofuranosidase;
XX KW genetic marker; plasmid; transformation.
XX OS Candida utilis.
XX FN WO9814600-A1.
XX PD 09-APR-1998.
XX PF 03-OCT-1997; 97WO-CU000005.
XX PR 03-OCT-1996; 96CU-00000082.
XX PA (INGG-) CENT ING GENETICA & BIOTECNOLOGIA.
XX PI Besabe Tuero L, Chavez Espinoza FP, Delgado Boada JM;
XX PI Gonzalez Martinez ME, Paifer Reyes E, Rivero Baeza T;
```

```
PI Rodriguez Menocal L;
XX WPI; 1998-240099/21..
XX N-PSDB; AAV19464.
XX PT New Candida utilis auxotrophic mutants and DNA markers - useful in
XX transformation systems to allow selection of transformants, e.g. for
XX heterologous protein production
XX PS Disclosure; Page 29-31; 63pp; Spanish.
XX CC This is the amino acid sequence of the Candida utilis INV1 anzyme
XX invertase (beta-fructofuranosidase). The invention relates to novel
XX C. utilis strains lacking one of the URA3 (AAV19462), HIS3 (AAV19463) or
XX INV1 genes resulting in auxotrophic mutants. The genes can be used
XX as markers in plasmids for transforming C. utilis. This organism has
XX not been used as a host for transformation due to difficulties in
XX obtaining reliable markers. Straightforward selection of transformants
XX using the new auxotrophic mutants as host cells is now possible.
XX SQ Sequence 533 AA;
Query Match 14.6%; Score 416.5; DB 19; Length 533;
Best Local Similarity 25.9%; Pred. No. 1e-27;
Matches 132; Conservative 87; Mismatches 182; Indels 109; Gaps 22;
Qy 40 RAVYHMTPPSGWLCDPQRPV--TTHGAYQLYLHSDON---NGPGWDHASTTDCGAVFTH 94
Db 30 RPLVHLTPPVGMNDPNGLFYDSSSTHYVYQYNNPNTINGCLPLYWGHATSDLLTWDH 89
Qy 95 HG--TYMPLRDPFVWSGSVAVGTANTAGFAGAGA---VVALATOPTDGVKRYQEOYL 147
Db 90 HAPALGPNDDGEGYSGSIVDIDYDNTSGFFDSDRPEQRIVAIYTNLPLDV---ETQDIA 146
Qy 148 WSTDGCGFTTALPDVIVNTDGRAATTAEIENAEWFRDPKIHWDTFARGEWVCVIGRLR- 206
Db 147 YSTDGCGYTFEYENNPVIDV-----NSTQFRDPKVIWYEETEQWVMTVAKSQE 194
Qy 207 -YAAFTSPNLRDWTLLRRNFDYPNHALGIECPDLFEITAD-----DGRHVLVAASMDA 260
Db 195 YKIQIYTSNLDKNSLASNFSTKGVGYQYECPLFEATIEPNKSGDPEKKWVWLAINP 254
Qy 261 YGIGLPM---TYAYWTGTWGDGEQFHADDLTPOWLQNGWDWYAAVTPSIDAPETRLRLA 317
Db 255 ---GSPLGGSINEXFVGDFNGTEFTPDDDATRFMDTKDFYAFQAF--FNAPENRSIGVA 309
Qy 318 WNNNKKYAAARDVPTDASDGYNGQNSIVRE-----LRLARQPGGWYTLTSTP 363
Db 310 WSSNMQY-SNQVPD--PDGYRSSMSSIREYTLRYVYSTNPESQILILCQK---FFVNETD 363
Qy 364 VAALTNYVTATTTLPDRTVD--GSAVLPMNGRAYEIELDIAWDTATNVGISVGRSPDGT 420
Db 364 LKVEEYKVSNSL---TVDHFTGSSFANSNTTGL---LDP-----NMTFTVNGTTDVT 411
Qy 421 RHTNIGKYGADLYVDRGPSDLA---GYSL-----APYSR 451
Db 412 QKDSV---TEELRIKSNQSDAIALGYDYNNEQFYINRATESYFQRTNQFOERSTVYVQ 468
Qy 452 AAAPIDPGARSVHLRLILYDTSQSVFVFNAG 481
Db 469 PLTITESGDKQYQLYGLVDNINILELYFNDG 498
RESULT 11
AAR72566
ID AAR72566 standard; Protein; 490 AA.
XX AC AAR72566;
XX DT 23-NOV-1995 (first entry)
XX DE Penicillium purpurogenum endo-inulinase.
XX XX
```

Penicillium purpogenum variety rubrisclerotium; endo-inulinase;
inulo-oligosaccharides; inulin.
Penicillium purpogenum.
JP07059574-A.
07-MAR-1995.
23-AUG-1993; 93JP-0229448.
23-AUG-1993; 93JP-0229448.
(MITK) MITSUI TOATSU CHEM INC.
WPI; 1995-135898/18.
N-PSDB; AAQ86692.

A new DNA encoding endo-inulinase - useful for the production of
inulo-oligosaccharides

Claim 1; Pages 4-5; 8pp; Japanese.

AAQ86692 encodes ARR72566 the Penicillium purpogenum variety
rubrisclerotium endo-inulinase. The endo-inulinase can be used
to prepare an inulo-oligosaccharide from inulin.

Sequence' 490 AA;

Query Match| 13.78; Score 389.5; DB 16; Length 490;
Best Local Similarity 26.08; Pred. No. 2.1e-25;
Matches 133; Conservative 87; Mismatches 225; Indels 67; Gaps 19;

QY 40 RAVVHTMPSPSGWLCDPQPVTTHGAYOLYLHSDNN--GPGWDHASTTIDGVAFTHGT 97
| : | : | : | : | : | : | : | : | : | : | : | : | : |
D b 4 RPTFHCPAENNMNEPNGLIKIDSTWHLYOADPTANWMGNMGCHATSSDLLHWDLHPV 63
| : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 98 VMPLRPDPFPPVMGSNAVVGTTANTAGFGAGA---VVALATQPTDGVRKYEOXYLWSTDGGF 154
| : | : | : | : | : | : | : | : | : | : | : | : | : |
D b 64 AIPVENGESFTGSYVDANNNTSSLTGSTNPYPYLAFFGTGYTSS-NGTDQRLAYSTDLGT 122
| : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 155 TFTAALP-DPVIVTDGRAATTPAEIEANEWRPKIHHDTARGEWCVI--GRURYAAFY 211
| : | : | : | : | : | : | : | : | : | : | : | : | : |
D b 123 TWLFESGNPII----SAALEAPHDVITGLESRDPKVFHEPSGRKMVMYLAHGQQDKLTFW 178
| : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 212 TSNLRDWTLRRNF-----DYNPHALGGICEPDLEITADDGTRH--WVL---AASMD 259
| : | : | : | : | : | : | : | : | : | : | : | : | : |
D b 179 TSLSAKSWTWMSDDLASIQEGFPS-SVWGVEVPDMFQLPI-OQNETHWTWIIFTPAQOSP 236
| : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 260 AYIGLPMTYA YMTGDGEQFHAD--DLTPQWLDGWDMWYAATWPSIDAPETKRLLAI 317
| : | : | : | : | : | : | : | : | : | : | : | : | : |
D b 237 AGGNV---VALTGSFDGETFLANPVSSTLLDYDRDFDGMASWENVPASDRLIIAA 292
| : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 318 WMNNKKYAARDVPDDASDGYNQNSIVRELRLARPQGWTLLSTPVAALTNYVTATTTL 377
| : | : | : | : | : | : | : | : | : | : | : | : | : |
D b 293 VMNSY---GSNPPTNT---WKGMLSFPRTLTLEIGSKQY-FLOOPTAELSTVDNALASI 345
| : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 378 PDRTVDGS AVL PWNGRAYEIELDTAWDTATNVGISVGRSPDTRTN I --CKYGADLYVD 435
| : | : | : | : | : | : | : | : | : | : | : | : | : |
D b 346 QNQTIA PKQTL LLSIHGSSLDRI AF SVDSGATULSLAVRKGGSEOTVI RYSQSNSTLSVD 405
| : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 436 RGPSDLAGYSLAPYRAAAP IDPCARSVH-----LRLIVDTOSQSVFVNAGH 482
| : | : | : | : | : | : | : | : | : | : | : | : | : |
D b 406 RTASGDISY-----DPAAGSIHASQA LARDNTELVYLRVL VDTCSVEVFGGGE 453
| : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 483 TVLSQQVHFHAEGDTGISLYTDGGPAHFTGI VV 514
| : | : | : | : | : | : | : | : | : | : | : | : | : |
D b 454 AVISDLIFPSSNDGLSLEVIGGTATLQSV EV 485
| : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 12
AADY25346
ID RAY25346 standard; Protein: 750 AA.

QY		318	WMNNKYYAARDVPTDASGYNQNSIVRE-----LRLAROPGGWYTLLSTPVAALT	368
Db		675	WMTNNDYPF-SFFT--TGWKGLSIPIRQVSLKETEEGRHMHTPIBELAQLRSPVLHIT	730
QY		369	N 369	
Db		731	N 731	
RESULT 14		AARS6256		
ID		AARS6256	standard; Protein; 556 AA.	
XX		AC	AARS6256;	
XX		DT	17-FEB-1995 (first entry)	
XX		DE	Inulinase preprotein.	
KW		Polymerase chain reaction; primer; amplify; PCR; extension assay;		
KW		functionality; inulinase; promoter; regulatory region; promoter;		
KW		upstream activating sequence; upstream repressing sequence; UAS; URS;		
KW		terminator; secretory signal; secretion; yeast; transformation;		
OS		Kluyveromyces marxianus.		
FH		Key Location/Qualifiers		
FT		Peptide 1..16 /note= "Pre-sequence"		
FT		Binding-site 13..19 /note= "Recognition site for signal peptidase"		
FT		Cleavage-site 16..17 /note= "Cleavage site for signal peptidase"		
FT		Protein 17..556 /note= "Mature Inulinase"		
FT		Binding-site 22..23 /note= "Processing site for KEX-2 like endoprotease"		
FT		Cleavage-site 23..24 /note= "Cleavage site for KEX-2 like endoprotease"		
PN		WO9413821-A.		
PD		23-JUN-1994.		
PF		09-DEC-1993; 93WO-EP03547.		
PR		11-DEC-1992; 92EP-0203932.		
PR		29-JAN-1993; 93EP-0200240.		
PA		(UNIL) QUEST INT BV.		
PI		Chapman JW, Musters W, Rouwenhorst RJ, Toschka HY, Verbakel JMA;		
DR		WI; 1994-217894/26.		
DR		N-PSDB; AAO66475.		
XX		Yeast regulatory sequence from the inulinase gene - for protein,		
XX		RNA or metabolite prodn.		
PS		Disclosure; Fig 5; 68pp; English.		
CC		This sequence represents the inulinase enzyme from K. marxianus. The		
CC		DNA encoding this protein was obtained by using the two primer sequences		
CC		given in AAO66460-61 to amplify a fragment corresponding to the		
CC		inulinase 5' region and using the amplified fragment as a probe in the		
CC		isolation of the full length sequence. In sequence analysis of the		
CC		isolated gene, three dominant signals emerged coinciding with T-174,		
CC		C-170 and C-167. These nucleotides are located approx. 100 nucleotides		
CC		behind the TATA box. Association of the start of transcription with this		
CC		region defines an uncommonly long 5' non-coding region. The MIGI binding		
CC		sequence perfectly matches the MIGI binding site consensus sequence from		
Db		731 N 731		
RESULT 13		AAAY24924		
ID		AAAY24924	standard; Protein; 750 AA.	
XX		AC	AAAY24924;	
XX		DT	26-AUG-1999 (first entry)	
XX		DE	Bacillus L7 endo-levanase.	
XX		Bacillus L7; endo-levanase; levan removal; sugar juice; sucrase;		
KW		invertase; dextranase; filtration.		
OS		Bacillus L7.		
XX		WO9931281-Al.		
XX		24-JUN-1999.		
XX		16-DEC-1998; 98WO-FI00988.		
XX		16-DEC-1997; 97US-0069801.		
PR		16-DEC-1997; 97US-0069800.		
XX		(CULT-) CULTOR CORP.		
PA		Miasnikov A;		
PI		WTI; 1999-395193/33.		
DR		N-PSDB; AAX83630.		
XX		Improving the filterability of sugar extraction juice with an		
PT		endo-levanase enzyme		
PS		Example; Page 19-22; 30pp; English.		
XX		The present invention describes a method for improving the filterability		
CC		of sugar extraction juice using an endo-levanase enzyme preparation		
CC		devoid of sucrose (invertase) activity. The endo-levanase enzyme		
CC		preparation is used (in combination with a dextranase) for treating the		
CC		sugar extraction juice, specifically to remove levan. The selectivity		
CC		of the endo-levanase enzyme solves the levan associated problems in the		
CC		beet sugar industry. The present sequence represents Bacillus L7		
CC		endo-levanase, for use in the method of the invention.		
Sequence		750 AA;		
Query Match		13.6%; Score 386; DB 20; Length 750;		
Best Local Similarity		31.9%; Pred. No. 8e-25;		
Matches		115; Conservative 44; Mismatches 144; Indels 58; Gaps 16;		
QY		40 RAVYHTPPSGWLCDPQPVTTHGAYQLYLHSQNGPGWDHASTTDGVAFTHGTVM	99	
Db		398 RPQYHYSPIRGSSADPNGLVFEGEYHLF--HQD----GGQWAHAWSROLINHWKLPLAL	451	
QY		100 PLRPDPFVMSGSNAVCTANTAGF-----GAGAVVALATOPTDVRYKYQOYLWSTDGCF	154	
Db		452 PWNLDGHVMSGSNAVADTTNASLGFGSGKGFLTAYTTSNPDRHNQNKIGLAYSTDGR	511	
QY		155 TFTALPD-PVVNTDCGRAATTAEIAENAEW-FRDPKIHMDTAGEWVCVIGRLRVAAFYT	212	
Db		512 TWKYSEHPVIENPKTGEDP-----GWDFRDPKVVRDEANNRWVVSOGDHIRLET	566	
QY		213 SPNLRTMLRRFPDYFNHALGGT-ECPDLFEITADDG-TRHVLAASMDAYGIGLPMTYA	270	
Db		567 STNLLNWTLTDQGYGAYIRGGWECPDLFOLPVESKKRKVKLMISTGAN---PNTQG	622	
QY		271 YWTGTWGEQGFHADLTLPQH-----LWGWDWYAAVWPISDAETKRRLAFA	317	
Db		623 S-----DAEYF-IIGDILTPECKFINDNPACTLTKTWGKEYYASMGSF--DMDPDGRIMIA	674	

CC the *S. cerevisiae* SUC2, GAL4 and GAL1 genes. The inulinase gene appears
CC to be regulated solely by glucose repression which allows the
CC construction of a strong, non-repressible promoter by exchanging the
CC Mig1 DNA binding site in the inulinase promoter. The regulatory region
CC of the inulinase gene comprises a promoter, an upstream activating or
CC repressing sequence (UAS or URS), a terminator as a regulatory region,
CC and a sequence encoding a secretory signal necessary for secreting a
CC gene product from a yeast. This regulatory region may be used to
CC prepare transformed yeasts for the production of desired gene products,
CC eg. protein or metabolites.

XX Sequence 556 AA;

Query Match 12.9%; Score 366; DB 15; Length 556;
Best Local Similarity 27.0%; Pred. No. 2.9e-23;
Matches 132; Conservative 72; Mismatches 181; Indels 104; Gaps 24;

QY 40 RAVYHMTTPSGWCLDQRPVTHGA-----YQLYLHSDN-----NGPGGDHASTTD 88

Db 39 RPSVHTPSHGWNNDP-----NGLWDAKEEDHLLYQYNPAATWTGTPLYMGHVSVD 92

QY 89 GVAFTHHGTVM-PLRDPFVWSSAVVGTANTAGF-----GAGAVVALATOPTDGV 138

Db 93 LRSWTYDASLPGSDDAGAFSGSMVIDYNTSGFFNSVDPQRQAVAVWTLKSGPS--- 149

QY 139 RYQEQYLYWSTDGFTTALPDPPVIVNTDGRAATTPAEIENAEWPRDPKLIHWDIAR--- 195

Db 150 ---QAQHISYSLDGGYTFQHYSDNAVLDI-----NSSNRPDPKVFHGEENCE 194

QY 196 -GENWCVICRLRY--AAFTSPNLRDWTLRNFDYPNHALGIECPDL-----FEITADDG 248

Db 195 DGRWIMAVAESQVSVLYFSSPNLKNLTLESNFTTHGWTGTQYECPLVKVPYDSVADSS 254

QY 249 TRH-----WVLAASMDAYG-IGLPMYAYWTGTDGEOFHADDLTPOWLDWGDWYA 299

Db 255 SNSSDSKPSDANWLVFVSNPGPLGSGVT-QYFVGDFNGTHETPIDDQTRFLDMGKDYA 313

QY 300 AVTWPSIDAPETKRL-AIAWMNNKYAARDVPTDASDGYNGQNSIVRELRL---ARQPGG 355

Db 314 LQTF--FNTPEKNDYGVIAWASWQY-AQAAPTDP---WRSSMSLVRFQTLKDFSTNPS 367

QY 356 WYTEL-STPVAALTNVYATTLTPDRTVDGSAVLPHNGRAYELEDIANDTATNVGISVG 414

Db 368 ADVVLNSQPV---LNYDALRKNKTTVITNVTSENGK--KIKLD----- 408

QY 415 RSPDGRTHRNICKYGADLYVDRGPSDLAGYSLAPYSRAAAPDPGARSVHLRLVDTQSV 474

Db 409 -NPGSGLEFHE-----YVFNQSPDIKSNVFDLSLYFKGNDD---NEYLRLGYETNGG 459

QY 475 EVFVNAGHT 483

Db 460 AFFLDRGHT 468

RESULT 15

AAP40020

ID AAP40020 standard; Protein; 532 AA.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX

PR

XX

PA

XX

PI

XX

XX

DR

DR

XX

PT

XX

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

25-APR-1983; 83US-0488337.

(GETH) GENENTECH INC.

Chang CN, Matteucci MD, Hitzeman RA;

WPI; 1984-301996/49.

N-PSDB; AAN40009.

Yeast expression vehicle - contains DNA sequence of yeast promoter
linked to DNA sequence encoding homologous protein.

Disclosure; Fig. 13; 53pp; English.

The DNA encoding the signal peptide of *S. cerevisiae* invertase may
be used in a new yeast expression vector for the expression of, esp.
human leukocyte interferon-alpha. The expression vector comprises a
yeast promoter operably linked to the signal sequence. This is attached
to the coding sequence of the protein. Yeasts transformed with this
vector produce and secrete protein normally heterologous to them.
The protein can be recovered from the medium free of presequences
etc. Human leukocyte interferon can be used as an antiviral or
antitumour agent. The promoter is esp. the invertase promoter or
the yeast 3-phosphoglycerate kinase promoter.

Sequence 532 AA;

Query Match 12.6%; Score 359.5; DB 5; Length 532;

Best Local Similarity 26.0%; Pred. No. 1e-22;
Matches 143; Conservative 81; Mismatches 228; Indels 99; Gaps 25;

QY 19 LALIFGAVPPAASAPGSLRAVYHMTTPSGWCLDP-----QRPVTHGAYQLYLHS 72

Db 7 LFLAGFAAKISASMTNETSDRLVHFTPKGWMNDPGLWYDEKDAKWH---LYFOYN 62

QY 73 DON---NGPGGDHASTTDCGVAFTTHG-TVMPLRDPFVWSSAVVGTANTAGFAGAV- 127

Db 63 PNDTVMGTLFLFGHATSDTLTNWEDQPIAIAKRNDSGAFSGSMVVDYNTSGFFNDTID 122

QY 128 ----VALATQPTDGVKRYQEQYLYWSTDGFTFTAL-PDPVIVNTDGRAATTPAEIENA 181

Db 123 PRQRCVAVITNTP---ESEEQIYSLOGGYTFTEYQKNPVL-----ANS 166

QY 182 EWRDPKIHWDTARGEVWCVIGRLR--YAAFYTPNLRDWTLRNFDYPNHALGG--IEC 237

Db 167 TQFROPKVFWEPSQKWIIMTAASKSDYKTEIYSSDDLKSKWLESAP--ANEGFLGYQEC 224

QY 238 PDLFEITA--DGTRHVLAAASMDAYGIGLPMYTA---YWTGTDGEOQHADDLTPOWLD 292

Db 225 PGLIEVPTQDPSKSYWVMSINP---GAPAGGSFNOYFVGSFNGTHEAFDQSRVVD 281

QY 293 WGDWYAAVTPSIDAPETKRLAIAWMNNKYAARDVPTDASDGYNGQNSIVRELRL--- 349

Db 282 FGKDIYALQTFNTDPTGYSGALGIAWASWEYSA-FVPTNP---WRSSMSLVKRFSLNTE 337

QY 350 --AROPGGWYTLSTPVAALTN-----YVATTTTLTPDRT--VDGSAVLFWNGRAYEIEL 399

Db 338 YOANPETELINKAEPILNISAGPWSRATNTTLTKANSYND---LSNSTGTLEFEL 393

QY 400 DTAWDTATNVGISVG-----RSPDGRTHITNGKYCADLYVDGRGSDLAGYSLAPY 449

Db 394 VYAVNTQTISKVFADLSLWFKGLEDPPEYLRMGFEVSASSFFFLDRGSKVKFKENPY 453

QY 450 --SRAAAPIDPGARSVHLRI-----LVDTSQSEVFEVFNAGHTVLSQCVHFAEGD----- 495

Db 454 FTRNRSVNNQPKSENDLSYKVGGLDQNIILELYFNDGDVYSTNTYFTMTGNALGSVM 513

QY 496 -TCIS--LYTD 503

Db 514 TTGVDNLFYID 524

Search completed: May 6, 2003, 12:32:12
Job time : 39.0673 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:30:07 ; Search time 29.0558 seconds
(without alignments)
1723.790 Million cell updates/sec

Title: US-09-868-328B-1
Perfect score: 2845
Sequence: 1 MTPAISRRVQLQGAGAGALA.....TDGGPAFTGIWREIGQAI 521

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2355	82.8	517	JC5875	levansucrase (EC 2
2	944	33.2	516	E70035	levanase homolog y
3	880.5	30.9	514	G97118	levanase (imported
4	562.5	19.8	512	A52006	sucrose/fructan hy
5	556	19.5	677	A27286	levanase (EC 3.2.1
6	515	18.1	622	A36915	fructanase - Bacte
7	467.5	16.4	943	JC4081	sucrase/fructanase
8	434	15.3	516	JEO301	inulinase (EC 3.2.
9	425	14.9	515	JC7533	inulinase (EC 3.2.
10	406.5	14.3	1142	C97080	levanase/invertase
11	389.5	13.7	515	JC5458	inulinase (EC 3.2.
12	382.5	13.4	1423	A49206	exo-beta-D-fructos
13	379	13.3	550	S65753	beta-fructofuranos
14	369.5	13.0	532	S27372	beta-fructofuranos
15	366.5	12.9	532	S27373	beta-fructofuranos
16	366	12.9	556	S31330	inulinase (EC 3.2.
17	365.5	12.8	555	S17502	inulinase (EC 3.2.
18	359.5	12.6	532	IFBY	beta-fructofuranos
19	355	12.5	477	C91034	sucrose-6-phosphat
20	354	12.4	477	S52162	sucrose hydrolase
21	348	12.2	477	D85878	sucrose hydrolase
22	348	12.2	581	JE80102	invertase - fissio
23	330	11.6	476	C34317	beta-fructofuranos
24	320	11.2	533	S13528	beta-fructofuranos
25	318	11.2	490	T43149	probable beta-fruc
26	313.5	11.0	453	T39155	glycosyl hydrolase
27	312	11.0	590	T02096	probable beta-fruc
28	297.5	10.5	581	T12095	beta-fructofuranos
29	283.5	10.0	580	S49308	beta-fructofuranos

30	283	9.9	650	2	S23217	beta-fructofuranos
31	276	9.7	512	2	JU0460	beta-fructofuranos
32	274.5	9.6	636	1	S31157	beta-fructofuranos
33	274.5	9.6	636	1	S31155	beta-fructofuranos
34	272	9.6	479	2	S68598	sucrose-6-phosphat
35	262	9.2	584	2	T06163	beta-fructofuranos
36	261.5	9.2	582	2	S37047	beta-fructofuranos
37	260.5	9.2	640	2	T09534	probable beta-fruc
38	257	9.0	661	2	S37590	beta-fructofuranos
39	256	9.0	592	2	S61503	beta-fructofuranos
40	256	9.0	661	2	S37592	beta-fructofuranos
41	255	9.0	634	2	S31925	beta-fructofuranos
42	253.5	8.9	562	2	S57951	beta-fructofuranos
43	253	8.9	587	2	S36231	beta-fructofuranos
44	250.5	8.8	639	2	S71268	beta-fructofuranos
45	249	8.8	642	2	S55521	beta-fructofuranos

ALIGNMENTS

RESULT 1

JC5875
levansucrase (EC 2.4.1.10) precursor - Arthrobacter nicotinovorans
N:Alternate names: levan fructotransferase
C:Species: Arthrobacter nicotinovorans
C:Date: 09-Mar-1998 #sequence_revision 09-Mar-1998 #text_change 20-Jun-2000
C:Accession: JC5875; PC4508
R:Saito, K.; Yokota, A.; Tomita, F.
Biosci. Biotechnol. Biochem. 61, 2076-2079, 1997
A:Title: Molecular cloning of levan fructotransferase gene from Arthrobacter nicotino
A:Reference number: JC5875; MUID:98101974; PMID:9438987
A:Accession: JC5875
A:Molecule type: DNA
A:Residues: 1-517 <SA1>
A:Cross-references: DBJ:AB001984; NID:92897799; PIDN:BAA24915.1; PID:92897800
A:Experimental source: strain GS-9
A:Accession: PC4508
A:Molecule type: protein
A:Residues: 34-45;137-166;185-238;309-319;424-445 <SA2>
C:Comment: This enzyme is involved in the production of levan which is a beta-2,6-li
C:Genetics:
A:Gene: lft
C:Superfamily: Penicillium purpogenum inulinase
C:Keywords: glycosyltransferase; hexosyltransferase
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-517/Product: levansucrase #status experimental <MAT>
F:52,232/Active site: Asp, Glu #status predicted

Query Match 82.8%; Score 2355; DB 2; Length 517;
Best Local Similarity 81.0%; Pred. No. 2.7e-166;
Matches 422; Conservative 41; Mismatches 54; Indels 4; Gaps 2;

QY	1	MTPAISRRVQLQGAGALALIFGGAVPPAARASAPGSLRAVYHMTTPPGSWLCLDPPRPVT	60
Db	1	MYDIDSRRLAQAGVGCALALEFMSNAIPVAAHAQA--SLRAIYHMTTPPGSWLCLDPPRPVH	58
QY	61	THGAYOLYLIHQSDNPGGWDHASTTDGVAFTHGFTVMPDRDPFVWGSAGVVGTA	120
Db	59	TNGAYOLYLIHQSDNPGGWDHASTTDGVSFTHGFTVMPDRDPFVWGSAGVVGTA	118
QY	121	GFAGAVVALATQPTDGVKRYQEQYLYWSTDGGFTFTALPDPIVNTDGRAATPAEIN	180
Db	119	GFAGAVIALATQPTDG--KFQEQYLYWSTDGGYSFTALPDPIVNTDGRATTAPEVEN	176
QY	181	AEWFRDPKTHWDATGECWVCVGRLYAAFYSPNLRDWTLRNFDYPNHALGGIECPDL	240
Db	177	AEWFRDPKTHWDATNRCWVCVGRLYAAFYSPNLRDWTLRNFDYPNHALGGIECPDL	236
QY	241	FETADGGTRHWVLAASMDAYGIGLPMYATWTGTWDGEQFHADDLTPQWLQWGDWYAA	300
Db	237	FEMTAGDGRHWVFGASMDAYSIGLPMYATWTGTWSWNGTAFADNLTPQWLQWGDWYAA	296

QY 301 VTPSIDAPETKRLAIAMNNWYKAAARDVPTDASDGYNGONSIVRELRLARQPGGWTTLL 360
Db 297 VTPPAVEAPETKRLATAMNNWYKAAARNVPTDASDGYNGONSITRELRLERQSGGWTTLL 356
QY 361 STPVAALTNVYATTTLPDRTRVGSAYLPPNGRAYETELDIANDTATNVGSGRSPDGT 420
Db 357 STPVALSNATSTTLPDRTRVGSAYLPPNGRAYETELDIANDTATNVGSGRSPDGT 416
QY 421 RHTNIGRYGADLYVDRGSLAGYSLAPYSRAAAPIDPGARSVHLRLVLTQSVFVNA 480
Db 417 RHTNIGRYGADLYVDRGSLAGYSLAPYSRAAAPIDPGARSVHLRLVLTQSVFVNA 476
QY 481 GHTVLSQVHFAGDGTGSLYTDGGAHFTGIVVREIGQAI 521
Db 477 GHTVLSQVHFAGDGTGSLYTDGGAHFTGIVVREIGQAI 517

RESULT 2

E70035
levanase homolog yveB - Bacillus subtilis

C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C:Accession: E70035
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertoni, C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chao, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997

A:Authors: Frenz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Ghimwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, A.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, H.; Masuda, S.; Maueel, C.; Medigu, R.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Portetelie, D.; Porwol, A.; Authors: Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Scanlon, S.J.; Schleier, amakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, A.; Authors: Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377

C:Accession: E70035

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
C:Genetics:

A:Molecule type: DNA

A:Residues: 1-516 <KUN>

A:Cross-references: GB:299121; GB:AL009126; NID:92635827; PIDN:CABL15451.1; PID:g2635959

A:Experimental source: strain 168

C:Genetics:

A:Gene: yveB

C:Superfamily: Penicillium purporogenum inulinase

Query Match 33.2%; Score 944; DB 2; Length 516;

Best Local Similarity 41.8%; Pred. No. 4.9e-62;

Matches 205; Conservative 71; Mismatches 176; Indels 38; Gaps 15;

QY 40 RAYVHTPPSGWGLCDPQRPVTHGAYOLYXLS-DOINGPG-CWDHASTTGDVAFTHHGT 97
Db 44 RAYVHTPPSGWGLCDPQRPVTHGAYOLYXLS-DOINGPG-CWDHASTTGDVAFTHHGT 103
QY 98 VMP--LRPDPFVWSSGAVVGTANTAGFAGAVVALATQPTDGVKRYQEOYLYWSTDDGFT 155
Db 104 AIPKYNPDGDIWGSVVDKENTAGFGKALVAIVTQPSAKDKK-QEOYLYWSTDDGFT 162
QY 156 F-----TALPDVIVNTDGRAATTPAEIENAEFRDPKIHWDTARGEVVCVIGRLRYAA 209
Db 163 FKPSYGNVMPNP---GTDD-----FRDPKIIWDFKNNKVMVLABGTGKIGFVSY 205
QY 210 FYTSPNLRDWTLRNFDPNHALGGIECPDLFEITADDGTRHVLAAASMDAYGIGLPMY 269
Db 206 FYSDNLKMDHYTSGF-FPEQA-GMVECPDLYNMARSDDGTRHVLAAASMDAYGIGLPMY 263
QY 270 AYTGTWDEQFADHDLTPQWLDGWDMWYAAVTPWSIDA--PETKRLAIAMNNWYKAAAR 327
Db 264 AYTGTWDEQFADHDLTPQWLDGWDMWYAAVTPWSIDA--PETKRLAIAMNNWYKAAAR 322

QY 328 DVPTDASDGYNGONSIVRELRLARQPGGWTTLLSTPVAALTNVYATTTLPDRTRVGSAY 387
Db 323 NTPT-MKNGFNGDTSVIRELRLKEQ-DGTYSLVSPQTEALEQLTSTDEIEDQDNGSKT 380
QY 388 LPWNGRAYETELDIANDTATNVGSGRSPDGTGTHHTNIGKY--GADLYVDRGSPDLGYS 445
Db 381 LSTIGDTQYQDTLWSLSKLNAGVRLRESBQDKRHIDVGFAGGYAYVNRATNQPKDS 440
QY 446 LAPYSRAAAPIDPGARSVHLRLVLTQSVFVNAAGTHTVLSQVHFAGDGTGSLYTDG 505
Db 441 -NTYVESKAPYDYNKRKVKHLKILVDKTTIEFVFGDGKTFVNEVPKPKDGITLYSDGG 499
QY 506 PAHFTGIVR 515
Db 500 TASFKNITVK 509

RESULT 3

G97118
levanase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001

C:Accession: G97118
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: G97118
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK79738.1; PID:g15024743; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC1773

C:Superfamily: Penicillium purporogenum inulinase

Query Match 30.9%; Score 880.5; DB 2; Length 514;

Best Local Similarity 40.7%; Pred. No. 2.4e-57;

Matches 200; Conservative 76; Mismatches 179; Indels 37; Gaps 14;

QY 40 RAYVHTPPSGWGLCDPQRPVTHGAYOLYXLS-DOINGPG-CWDHASTTGDVAFTHHGT 97
Db 45 REYVHTFVTPKNNKNDPQRPVDFDGEYHYLYNKDYPNGNGTEWROQATSKDLTVWDEGV 104
QY 98 VMP--LRPDPFVWSSGAVVGTANTAGFAGAVVALATQPTDGVKRYQEOYLYWSTDDGFT 155
Db 105 CIPKYNENGDIWGSFVVDQAQNTAGFGKALVAIVTQPSAKDKK-QEOYLYWSTDDGFT 163
QY 156 FTALPD-PVIVNTDGRAATTPAEIENAEFRDPKIHWDTARGEVVCVIGRLRYAAFTSP 214
Db 164 FKPSYGNVMPNP---GTDD-----FRDPKIIWDFKNNKVMVLABGTGKIGFVSY 211
QY 215 NLRDWTLRNFDPNHALGGIECPDLFEITADDGTRHVLAAASMDAYGIGLPMYAYWTG 274
Db 212 NLKNW--QHTGDFFTNIGIVECPDIFMOSDNGNTKWLGTSGANGKVSSEPTAYVWG 269
QY 275 TWDEQFADHDLTPQWLDGWDMWYAAVTPWSIDA--KRLAIAMNNWYKAAARVPTD 332
Db 270 NYDKKFIADISTPRLWDYGFDMWYAAVTPWSIDA--KRLAIAMNNWYKAAARVPTD 327
QY 333 ASDYNGONSIVRELRLARQPGGWTTLLSTPVAALTNVYATTTLPDRTRVGSAYLPMW 392
Db 328 IONRENGMSIVREITLSKQKNTYSLSKPIKRIENITTSIDQFQKISVKGLKHLKVG 387
QY 393 RAYEIELDANDTATNVGSGRSPDGTGTHHTNIGKY--GADLYVDRGSPDLGYS 448
Db 388 KAYQLDNDINWDKAVNGVRLRESLDRKIDVGFAGGYAYVNRATNQPKDS 441
QY 449 ---YSRAAAPIDPGARSVHLRLVLTQSVFVNAAGTHTVLSQVHFAGDGTGSLYTDG 505
Db 442 SKKIVESRAPFDLNNKVKHLRFVDKVSVEVFIDDKITYSNEVFPKPKDGITLYS 501

Db 258 MOYFVGDFGTHFKENPNKVLMTDGRDYAAVSDIPSTDSRRLLWGLWMSWQY-A 316
QY 327 RQVPTDASDGYNGQNSIVRELRLARQPGWYTLTSTPVAALTNVYVATTLLPDRTPVDGSA 386
Db 317 NDVPTSP---WRSATSPRELKLAFTGEG-VRVVQTPVKLETTIRGTSKKWKNLTISPAS 372
QY 387 ---VLPWNGRAYEI--ELDIAMDTATNVGIVSGRSPDGTGRTNICKYGADLYVDRGPSDL 441
Db 373 HNVLAGQSGDAYEINAEFKVSPGSAEFGFKVGTGENQFTKGVYDRRAKLEFVDRSESGN 432
QY 442 AGYSLA-PYSRAAPIDPGARSVHLRLVDTQSVSEVFNAGHTVLSQOVHFAEGDTGISL 500
Db 433 DTFNPAFTKETAFLKPVNGKVLRIFFVDRSSVEFNGDKQVITDIIPLDRSSKGLEL 492
QY 501 YTDGGPAHFTGIVVREI 517
Db 493 YAANG-----GVKVKSL 504

RESULT 6
A36915
fructanase --Bacteroides fragilis
C:Species: Bacteroides fragilis
C>Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 17-Mar-2000
C:Accession: A36915
R:Blatch, G.L.; Woods, D.R.
J. Bacteriol. 175, 3058-3066, 1993
A:Title: Molecular characterization of a fructanase produced by Bacteroides fragilis BF-1
A:Reference number: A36915; MUID:93259952; PMID:8491724
A:Contents: BF-1
A:Accession: A36915
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-622 <BLA>
A:Cross-references: GB:M83774; NID:g143970; PIDN:AAA22924.1; PID:g143972
A:Note: sequence extracted from NCBI backbone (NCBIN:131947, NCBIPI:131949)
C:Superfamily: beta-fructofuranosidase

Query Match 18.1%; Score 515; DB 2; Length 622;
Best Local Similarity 30.2%; Pred. No. 3.1e-30;
Matches 156; Conservative 77; Mismatches 195; Indels 88; Gaps 20;

QY 40 RAVYHMTTPSGWGLCDPQRPVTHGAYOLYLLHSDQNN--GPGGWDHASTTGDGVAFTHHGT 97
Db 133 RPLYHHTPLYGWMDANGLVYKDEYHLFYQYNPYGSWMGMMHGHVSVDLVHWEH--- 189
QY 98 VMLRPDPF-----VMSGSAVVGTAAGAGAVVALATQPTDGVRYKQEQYLYWSTD 151
Db 190 ---LEPALARDTLGHIFSGSSVVDANTAGYAGAIYAFYTSASD--KNGOIOCMAYSTD 244
QY 152 GGFFTTAL-PDPVIVNTDGRAATTPAEIENAEWRDPKIHWDTARGEVWCVIGRLRYAAF 210
Db 245 NGRTFTYKRNPLVTPDQ-----LKDPRDKVFWYAPDQKVVWVVSADKEMRF 293
QY 211 YTSNLRDWTLLRNFD-----YPNHALGIECPDLFEITADGTRH--WVLAASMDAYGI 263
Db 294 YSSENLEWYMSGWGGYGVQPSQ-----PECPDMVELPVDGNPDHKKWALIVNVN--- 345
QY 264 GLPMY-----AYWTGTWGEQFHADDL--TPQWLDWGDWDYAAVTPSIDAPEKRLA 315
Db 346 --PGCYFGSATQYFIDFGDGEKFCVCKNPETVKWLDWGDWHDYATVCFSTNG---DRTIA 400
QY 316 IAWNNKRYAARDVPTDASDGYNGQNSIVRELRLARQPGWYTLTSTPVAALTNVYVATT 375
Db 401 VPWMSNKHQY-ANIVPTPQ---FRSANALPRELSLYTQGDYI-MAAPVETKSLRKESR 455
QY 376 TLDPRTVDGS-----AVLPWNGRAYEIELDIAMDTATNVGIVSGRSPDGTGRTNICKYGAD 431
Db 456 EIPAFEGDAYHVDLSLSDNKGAYEIELELAAGSAEIMGLKLFNEKGENVDIVISLPEKK 515
QY 432 LYVDRGPSDLA--GYSLAPYSRAA-----APIDPCARSVHLRLV 469

Db 516 LVMDRTKSGIVDFGKDSAPHAEIAHRRKQNSINYVDDELGTGWAPVQK-AGNTYKLDIEV 574
QY 470 DTQSVSEVFNAGHTVLSQOVHFAEGDTGISLYTDCG 505
Db 575 DKCSVEIFLNGGKIATNLIPTTPYQMSFYSRGG 610

RESULT 7
JC4081
sucrase/fructanase precursor - Actinomyces naeslundii (strain T14V)
C:Species: Actinomyces naeslundii
C>Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 15-Oct-1999
C:Accession: JC4081
R:Norman, J.M.; Bunney, K.L.; Giffard, P.M.
Gene 152, 93-98, 1995
A:Title: Characterization of levJ, a sucrase/fructanase-encoding gene from Actinomyces
A:Reference number: JC4081; MUID:95129923; PMID:7828936
A:Accession: JC4081
A:Molecule type: DNA
A:Residues: 1-943 <NOR>
A:Cross-references: GB:U12274; NID:g515689; PIDN:AAA67876.1; PID:g515690
A:Note: The authors translated the codon ATG for residue 467 as Asn, TAC for residue
C:Genetics:
A:Gene: levJ
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-943/Product: sucrase/fructanase #status predicted <MAT>

Query Match 16.4%; Score 467.5; DB 2; Length 943;
Best Local Similarity 21.5%; Pred. No. 1.8e-26;
Matches 187; Conservative 83; Mismatches 217; Indels 381; Gaps 27;

QY 4 AISRAVLOGAGAGALALIEGG-----AVPPAARASAPGS-LRAYVHMTTPSGWGLCDPQRP 58
Db 19 AASTLASMAAAASQARPVSAAPNAPAAPATAPKADQGTGERWRPQSHVTPKQNMNDPGL 78
QY 59 VTHGATQLYLLHSDQNN--GPGGWDHASTTGDGVAFTHHGTVMPLRDPFVWSSSAVVG 116
Db 79 VYDGEYHMFYQYNPYGSWMGMMHGHVSVDLVHWEH--- 138
QY 117 ANTAGFCA---GAVVALATQPTDGVRYKQEQYLYWSTDGGFETAL--PDPVIVNTDGRA 171
Db 139 KNTSGLSGSPDNPAWVWTRADYVGN--QSQSLAYSTDGKGTWNLNNGDPVL----- 189
QY 172 ATTPAEIENAEWRDPKIHWDTARGEVWCVIGRL--RYAAPTYSNLRDWTLLRNFDYPN 229
Db 190 -----DIGSNE-FRDPKVFWDQASGRWTVVSHATEHRVSFYSPDLIHWTQSSFGGEG 243
QY 230 HALGGIECPDLFEITADGTRH--WVLA----- 255
Db 244 ITSAVWACPDFFPLPVDGSSQEQVKWLVTVVADSQAQYFVGSWDGTTFTTPEIPHYSGEG 303
QY 256 -----ASMDAYG----- 262
Db 304 TTLADFENGYAGWKADGAAGFSGSPATGDLPGHQKAYVDSFGSGDADTGLTSDFTVSS 363
QY 263 ----- 262
Db 364 SYINLRTAGGKHPYNQATGONGGRLLAGFDGSGWEGTVEGSAFAATPPQATPAQOPL 423
QY 263 -----ICLPMY-----AY----- 271
Db 424 VNHSAGLLNTYLDATGOGSDAPTGTATPTFTTIDSAIYNLNLMMGGNNPRPEGGADGS 483
QY 272 -----W-----TGTW----- 276
Db 484 RVSVELIVDGKVVRSATGRNLEELNQSHDVSCLKGSAQIVVTDATGCGWHILLDEV 543
QY 277 ----- 276
Db 544 RASDKKASPIADNTSVNLVVDGKVVASATGNNSGTLEWTSNMVAAKYGRKARLVIEDRNG 603
QY 277 -----DGEQFHADDLTPQWLDWGDWHDYAAVTPSIDAPEKRLIAWMN 320

Db 604 NAEDGHLMVDDILOLSTKAFSGADVPR-LDYGKDYAAVTDWNV--PNGKRYQVGMWS 660
QY 321 NKYAAARDVPTDASDGYNGQNSIVRELRLARQPGWYTLTSPVAALTNYVTA-----T 374
Db 661 NWAY-VRLDPTTT---WRTAMSTVREMGLTR-VNGKRLRTAQPVTALESRLTQDELIRKD 715
QY 375 TTLP-DRIVDGSAPLWPNNGRAYEIELDAMDATNVGVSGRSPDGRHTNTG--KYGAD 431
Db 716 TDIPVGETSLGKAA---QCTSLDLSVDLSPSASSFAGLVK--LDNGEQVTLIGYDSQAQK 770
QY 432 LVVDGRPSDLAGYSLAPYSRAAPDPGAR-SVHLRLVLDVOSVEFVNAGHTVLSQQVH 490
Db 771 LVVDTRHSGVTDSPKFPARSTAPUSPSKQGVHLRLIIVDAHSVEFAADGTPVITQTVY 830
QY 491 FAEGDTGISLYTDGGAHFTGIWVREIG 518
Db 831 PRQDATGVSLYAEAGCTAHLGSLSLWHLG 858

RESULT 8
JE0301
Inulinase (EC 3.2.1.7) - Aspergillus niger
C:Species: Aspergillus niger
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 07-May-1999
C:Accession: JE0301
R:Ohta, K.; Akimoto, H.; Matsuda, S.; Toshimitsu, D.; Nakamura, T.
Biosci. Biotechnol. Biochem. 62, 1731-1738, 1998
A:Title: Molecular cloning and sequence analysis of two endoinulinase genes from Aspergillus niger
A:Reference number: JE0301; MUID:99022189; PMID:9805373
A:Accession: JE0301
A:Molecule type: mRNA
A:Residues: 1-516 <OHT>
A:Cross-references: DBJ:AB012771
C:Superfamily: Penicillium purporogenun inulinase
C:Keywords: glycosidase; hydrolase

Query Match 15.3%; Score 434; DB 2; Length 516;
Best Local Similarity 27.1%; Pred. No. 2.3e-24;
Matches 144; Conservative 92; Mismatches 22; Indels 70; Gaps 19;

QY 24 CGAVPPAARASAPGSLRAYVHMTTPSGWCLDPQRPVTHGAYQLYLLHSDQNN--W 81
Db 16 GLMLPSQAQSN---DYRPSYHFTPDQYWNPNGLIKGTWHLFPQHNPNTANVWNGTCW 72
QY 82 DHAISTDGAFTTHGCTVPLRDPFVWGSAGVGTANTAGFAGA---VVALATOPTDGV 138
Db 73 GHATSTDLMHAKHAKFTADENGVEAFTGTAYDPNNTSGLSDANPPYLAFTGYTTS- 131
QY 139 RYQEOYLWSPDGGFTTALPDVIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGEW 198
Db 132 SOTQDORLAFSDNGATWTKFOGNIISTSQEA---PHDITGGLSRDPKVFHRSQGNW 188
QY 199 VCVI--GRLRYAAFTSPNLRDWTLRNFDYPNHALGGI-----ECPDLEIITADG 249
Db 189 INVLAHGODKLSFWTSADTINWTQW---DLKSTSLNGLSDDITGWEPDMFELPV-EG 245
QY 250 RH--WVL---AASMDAYGIGLPMYAYWGTGWDCQEQHAD--DLTPQWLDWGDWYAAV 301
Db 246 BETTWVVMWTPAEGSPAGNGVLAI---TGSFDCSEKSTADPVDASTWLDNGRDFDGNAL 301
QY 302 TWPSIDAPETKRLATAMNNKYYAARDVPTDASDGYNGQNSIVRELRLARQPGWYTL 361
Db 302 SWNVNPSADGRRIIAAVNSY---GSPNPTTT---WKGMLSPRTLSL-KKVGTOQHVFQ 354
QY 362 TPVAALTNYVATTTLPRDVTGSAVLPWNGRAYEIELDIW--DTATNVGISVGRSPDG 419
Db 355 QPITELDPTISLQTLANOTITPGQTLSSIRGTALDVRVAFYPDAGSVLSLAVRKASE 414
QY 420 TRHNTIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVH-----LR 466
Db 415 QTVINYTQSNATLSVDRFESGDISY-----DPAAGGVHTAKLEEDGCTGLVSIR 462
QY 467 ILVDTQSVEFVNAGHTVLSQQVHFAEGDTGISLYTDGGAHFTGIWVREI 517

Db 463 VLVDTCSEVFGQGEAVISDLIFFSDSDGLALEVTGCGNAVLQSDVDRSV 513

RESULT 9

JE07533
Inulinase (EC 3.2.1.7) - Penicillium sp. (strain TN-88)
N:Alternate names: 2,1-beta-D-fructan fructanohydrolase
C:Species: Penicillium sp. (strain TN-88)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: JC7533

R:Akimoto, H.; Kiyota, N.; Kushima, T.; Nakamura, T.; Ohta, K.
Biosci. Biotechnol. Biochem. 64, 2328-2335, 2000
A:Title: Molecular cloning and sequence analysis of an endoinulinase gene from Penicillium sp.
A:Reference number: JC7533; MUID:21036892; PMID:11193399
A:Accession: JC7533
A:Molecule type: mRNA

A:Residues: 1-515 <AKI>
A:Cross-references: DBJ:AB041337

A:Experimental source: strain TN-88, wild type

C:Comment: This enzyme, an acidic glycoprotein and an endo-acting inulinase as well, is a
aose as the main products.

C:Genetics:

A:Gene: inuG

C:Superfamily: Penicillium purporogenun inulinase

C:Keywords: glycoprotein; glycosidase; hydrolase

Query Match 14.9%; Score 425; DB 2; Length 515;

Best Local Similarity 27.6%; Pred. No. 1.1e-23;

Matches 142; Conservative 92; Mismatches 232; Indels 48; Gaps 21;

QY 31 ARASAPGSLRAYVHMTTPSGWCLDPQRPVTHGAYQLYLLHSDQNN--GPGWMDHASTD 88
Db 21 ARAVA-DDYRPAHFPCPAENWNEPNGLLIQINSTWHLFYQADPAANVWGCWGHATSSD 79

QY 89 GVAFTHHGTVMLRDPDPVWGSAGVGTANTAGFAGA---VVALATQDTPGVKRYQEOY 145

Db 80 LLHWDLHPVAIPVENGIESFTGTSYDSNNTSGLTSTNPPYLAFTGYTES-NKTQDQR 138

QY 146 LWKSTDGGTFTALP-DPVIIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGEWCVI-- 202

Db 139 LAYSTDGLGTWVKFAGNPDI---GAAQEPQDISGGLSRDPKVFHAPSCKWVWVLAH 194

QY 203 GLRLYAAFTSPNLRDWTLRNFD-----YPNHALGIECPDLFEITADGTRH--WVL 254

Db 195 GQDKLITFTSLDAKNWTVSDLSSSQIBGFPS-SITGWEVPMFOLPI-QGIKKTTWVL 252

QY 255 ----AASMDAYGIGLPMYAYWGTGWDCQEQHADLTPO--WLDGMDWYAAVTPSIDA 308

Db 253 IFTPAQGSPPGGNV---VALTGSFDTGFTVADPVPSTLWLDYGRDFGALSWEVPA 308

QY 309 PETKRLATAMNNKYYAARDVPTDASDGYNGQNSIVRELRLARQPGWYTLTSPVAALT 368

Db 309 SDGRILIAAVNSY---GSPNPTTT---WKGMLSPRTLAL-KQIGSKOYFLOQPVAEUS 361

QY 369 NVVTATTTLPDRTVDSAVL--PWNGRAYEIELDIWDTATNVGISVGRSPDGRHTNIG 426

Db 362 TIDGSLTSTQNTQITPNQTLSSIHGTSIDIRMAFVIDSGATLSLAVRKG--GSEQTVIR 419

QY 427 KY--GADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLVLDVOSVEFVNAGHT 483

Db 420 YFQSNSTLSVDRITASGDISTYDPAAGGVHTAQLAQDNTELHVALIDTCSVEFVGEGBA 479

QY 484 VLSQQVHFAEGDTGISLYTDGGAHFTGIWVREI 517

Db 480 VISDLIFFPSNDSGLSLEVSGGTAMLRSVNVSSV 513

RESULT 10

C97080

levanase/invertase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

Db 800 LAITPLKSAV 809

RESULT 11

JC5458

inulinase (EC 3.2.1.7) precursor - Penicillium purporogenum
N:Alternate names: 2,1-beta-D-fructan fructanohydrolase
C:Species: Penicillium purporogenum
C:Date: 17-Jun-1997 #sequence_revision 17-Sep-1997 #text_change 20-Jun-2000
C:Accession: JC5458; PC4326
R:Onodera, S.; Murakami, T.; Ito, H.; Mori, H.; Matsui, H.; Honma, M.; Chiba, S.; Shi
Biosci. Biotechnol. Biochem. 60, 1780-1785, 1996
A:Title: Molecular cloning and nucleotide sequences of cDNA and gene encoding endo-in-
F:26-515/Product: endo-inulinase #status predicted <MAT>
A:Reference number: JC5458; MUID:97141595; PMID:8987853
A:Accession: JC5458
A:Molecule type: DNA
A:Residues: 1-515 <ON0>
A:Cross-references: DDBJ:D84360; NID:g1805497; PIDN:BAI2321.1; PID:g1805498
A:Accession: PC4326
A:Molecule type: protein
A:Residues: 26-70;152-239;333-345;350-379 <ON02>
C:Comment: This enzyme hydrolyzes inulin and fructo-oligo-saccharides, but not sucros
C:Superfamily: Penicillium purporogenum inulinase
C:Keywords: glycoprotein; glycosidase; hydrolase
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-515/Product: endo-inulinase #status predicted <MAT>
F:108,109,132,246,424,489/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 13.7%; Score 389.5; DB 2; Length 515;
Best Local Similarity 26.0%; Pred. No. 4.5e-21;
Matches 133; Conservative 87; Mismatches 225; Indels 67; Gaps 19;

Qy 40 RAVYHMPSPGSLCDPQRPVTHGAYOLYLYHSDNN--GPGGWDHASTTGDGVAFTHHGT 97
Db 29 RPTFHCPAENWMEPNGLIKIDSTWILFQADPTANVNGNECWGHATSSDLLHWDHLPV 88
Qy 98 VMPLRPDPVWMSGAVVGTANTAGFGAGA---VVALATQPTDGVRYQEOYLXWSTDGGF 154
Db 89 AIPVENGIESFTGTSYYDANNTSSLGSTNPPYLAFPTGYTSS-NGTQDQRLAYSTD LGT 147
Qy 155 TFTAALP--DPVIVNTDGRAATTPEIAENAFRDPKTHWDTARGEWCVI--GRLRYAAFY 211
Db 148 TWLFKSGNPIT---SAALEAPHDVTGGLESRDPKVFHFEPSGKVMVLAHGQDGLTFW 203
Qy 212 TSPNLRDWTLLRRNF-----DYPNHALGGTECPDLPFETADCDTRH--WVL----AASMD 259
Db 204 TSLDAKSWTWMSDLLASQIEGFPS-SVTGNEVPDMPFOLPI-QGTNETTWIITPAQGSP 261
Qy 260 AYGIGLPMTYAYWTGWDGQFIAD--DLTPQWLDMGWDWYAAVTPSIDAPETKRLAIA 317
Db 262 AGGNV---VALTGSFQGETFLANPVDSSSTLWDYGRDFDGAMSWENVPASDGRLLIAA 317
Qy 318 WMNNWKYAARDVPTDASDGYNGQNSIVREILARQPCGWYTLTLLSTPVAALTNNVTATTL 377
Db 318 VMNSY---GSNPPTNT---WKGMLSPFRTLTLEKIGSKQY-FLQQTIAELSTVDNALASI 370
Qy 378 PDRVDSGAVLPWNGRAYEITLDTANDTATNVGISVGRSPDGTTRHNI--GKYGADLYVD 435
Db 371 QNQTIAPKQTLILLSIHGSSLDVRIATSVDSGATLSLAVRKGSGEQVIRYQSQNSTLSVD 430
Qy 436 RGPSDLAGYSLAPYSRAAAPIDPCARSVH-----LRLIVDTQSVFVFNAGH 482
Db 431 RTASGDLSY-----DPAAGIHSQAQLARDNTELVILRLVLDVTSVEVFGGGE 478
Qy 483 TVLSQQVHFHAEQGTGISLYTDGGPAHFTGIW 514
Db 479 AVISDLIFPSSDGLSLEVGIGTATLQSV 510

RESULT 12

A49206

exo-beta-D-fructosidase - Streptococcus mutans
C:Species: Streptococcus mutans

A: Accession: S25441
A: Molecule type: DNA
A: Residues: 1-74 <HO2>

A;Cross-references: EMBL:X07570

C;Genetics:

A;Gene: SUC1

A;Map position: 7R

C;Superfamily: beta-fructofuranosidase

C;Keywords: glycosidase; hydrolase

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-532/Product: beta-fructofuranosidase 1 #status predicted <MAT>

Query Match 13.0%; Score 369.5; DB 2; Length 532;
Best Local Similarity 25.1%; Pred. No. 1.4e-19;
Matches 140; Conservative 84; Mismatches 221; Indels 113; Gaps 25;

QY 19 LALIFGAVPPAARASAPGSLRAVYHMTTPSGWLCDPQR--PVTHGAYOLYLHSDQNN 76

DQ 7 LFLLAGFAAKISASMTNETSDRPLVHFTPNKGMNDPGLWDYDAKEGKWHLYFOYNPDNT 66

QY 77 G--PGGWDHASTTDCGVAFTHHG---TVMPRLRDPFVWSSAVVGTANTAGCAGAV-- 127

DQ 67 VWGLPLFWGHATSDD---LTHWQDEPVAIAIPKRDGSGVSDVIDYNTSGFNDTIDP 123

QY 128 ---VALATOPTDGVRRYQOYLWSTDCGFTFTAL-PDPVIVNTDGRAATTAEIENAE 182

DQ 124 RQRCVAIWTYNTP---ESEEQYISYLDGGYTFTEYQKNPVL- 167

QY 183 WFRDPKIHMDTARGEWCVCIGRLR--YAAFYTSPNLRDWTLRNFDYPNHALGG--IECP 238

DQ 168 QFRDPKVFWEPSKKIWTAAKSDYKIEIYSSDDLKSWKLESAP--ANEGFLGYQYEC 225

QY 239 DLFEITA--DDGTRHWVLAASMDAYGICLPMTYA---YWTGTWGDGEFHADDLTPQWLDW 293

DQ 226 GLIEVSEQDPKSHWVMTISINP---GAPAGGSFNOYFVGSNGHFEAFDQSRVYDF 282

QY 294 GWDYAAVTWPSIDAPETKRLAIAMNNWKYAADVPTDASDGYNGONSIVRELRLAROP 353

DQ 283 GKDYALQTFENTDPTGSGALGIAWASNWEYSA-FVP---SNPWRSSMSLVRFPSLNT- 337

QY 354 GGWYTLSTPVAALTNVYATTTLPRTVDGSAVLPNW-----GRAYEIELDIADW 403

DQ 338 ---YQANPETELINLKA-----EPILNITSSAGPWSREATNTTLTKANSYVNDLSNT 386

QY 404 DT-----ATNVGISVGRS-----PDGTRHTNIGKYCADLYVDRGPSDLA 442

DQ 387 GTLEFELVAVNTQTTSKSVFADLSLWFKGLEDPPEYLRMGFEVSASSFFLDGNSKVK 446

QY 443 GYSLAPY--SRAAAPIDPGARSVHLRI-----LVDTSQSVFVNAGHTVLSQQVHFAEGD 495

DQ 447 FVENPYFTNRMSVNNQPFKSENDLSYKYVYGLLDQNLILELYFNDGDVVSNTYFMTTGN 506

QY 496 -----TGIS--LYTD 503

DQ 507 ALGVSNNMTTGVNDLFLYID 524

RESULT 15

S27373

beta-fructofuranosidase (EC 3.2.1.26) 4 precursor - yeast (Saccharomyces cerevisiae)

N;Alternate names: invertase; saccharase

C;Species: Saccharomyces cerevisiae

C;Date: 23-Apr-1993 #sequence_revision 23-Apr-1993 #text_change 22-Jun-1999

C;Accession: S27373; S25442

R;Hohmann, S.

submitted to the EMBL Data Library, November 1988

A;Reference number: S27372

A;Accession: S27373

A;Molecule type: DNA

A;Residues: 1-532 <HOH>

A;Cross-references: EMBL:X07572; NID:g4572; PIDN:CAA30459.1; PID:g4573

Mol. Gen. Genet. 211, 446-454, 1988

A;Title: Structural analysis of the 5' regions of yeast SUC genes revealed analogous pal

A;Reference number: S25439; MUID:88216256; PMID:2835632

A;Accession: S25442

A;Molecule type: DNA

A;Residues: 1-74 <HOR2>

A;Cross-references: EMBL:X07572

A;Note: the authors translated the codon ATT for residue 7 as Asn

C;Genetics:

A;Gene: SUC4

C;Superfamily: beta-fructofuranosidase

C;Keywords: glycosidase; hydrolase

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-532/Product: beta-fructofuranosidase 4 #status predicted <MAT>

Query Match 12.9%; Score 366.5; DB 2; Length 532;
Best Local Similarity 26.2%; Pred. No. 2.4e-19;
Matches 143; Conservative 80; Mismatches 231; Indels 91; Gaps 24;

QY 21 LIFGAVPPAARASAPGSLRAVYHMTTPSGWLCDPQR--PVTHGAYOLYLHSDQNNG- 77

DQ 9 LLAGFAAKISALMTNETSDRPLVHFTPNKGMNDPGLWDYDAKEGKWHLYFOYNPDNTV 68

QY 78 --PGWDHASTTDCGVAFTHHG---TVMPRLRDPFVWSSAVVGTANTAGCAGAV---- 127

DQ 69 GLPLFWGHATSND---LTHWQDEPVAIAIPKRDGSGVSDVIDHNTSEFFNDTVDPRQ 125

QY 128 --VALATOPTDGVRRYQOYLWSTDCGFTFTAL-PDPVIVNTDGRAATTAEIENAEWF 184

DQ 126 RCVAIWTYNTP---ESEEQYISYLDGGYTFTEYQKNPVL- 169

QY 185 RDPKIHMDTARGEWCVCIGRLR--YAAFYTSPNLRDWTLRNFDYPNHALGG--IECPDL 240

DQ 170 RDPKVFWEPSKKIWTAAKSDYKIEIYSSDDLKSWKLESAP--ANEGFLGYQYECPL 227

QY 241 FEITA--DDGTRHWVLAASMDAYGICLPMTYA---YWTGTWGDGEFHADDLTPQWLDGW 295

DQ 228 IEVPTEDQPSKSHWVMTISINP---GAPAGGSFNOYFVGSNGHFEAFDQSRVYDFGK 284

QY 296 DWYAAVTWPSIDAPETKRLAIAMNNWKYAADVPTDASDGYNGONSIVRELRL-----A 350

DQ 285 DYVALQTFENTDPTGSGALGIAWASNWEYSA-FVPTNP---WRSSMSLVRFPSLNTYQA 340

QY 351 RQPGWYTLSTPVAALTN-----YVTATTTLPRTVDGSAVLPNWNGRAYEIELDIADWT 405

DQ 341 NPETELINLKAEPILNINISNAGPWLHPASNTL-TKANFSFVDSLNSGTLEFELVAVNT 399

QY 406 ATNVGISVG-----RSPDGTRHTNIGKYCADLYVDRGPSDLAGYSLAPY--SRAA 453

DQ 400 TQSVKSVFSDLSLWFKGLEDPPEYLRMGFEVSASSFFLDGNSKVKFVKENPYFTNRMS 459

QY 454 APIDPGARSVHLRI-----LVDTSQSVFVNAGHTVLSQQVHFAEGD-----TGIS- 499

DQ 460 VNNOPFKSENDLSYKYVYGLLDQNLILELYFNDGDVVSNTYFMTTGNALGSVNNMTTGVND 519

QY 500 -LYTD 503

DQ 520 LFYID 524

Search completed: May 6, 2003, 12:35:46

Job time: 33.0558 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:23:52 ; Search time 10.5202 Seconds
(without alignments)
2054.066 Million cell updates/sec

Title: US-09-868-328B-1

Perfect score: 2845

Sequence: 1 MTPAIRRAVLQAGAGALA.....TDGGPAHFTGIVREIGQAI 521

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	556	19.5	677	1	SACC_BACSU	P05656 bacillus su
2	382.5	13.4	1423	1	FRUA_STRMU	Q03174 streptococ
3	379	13.3	550	1	INV1_HANAN	P40912 hansenua a
4	369.5	13.0	532	1	INV1_YEAST	P10594 saccharomyc
5	366.5	12.9	532	1	INV4_YEAST	P10596 saccharomyc
6	365.5	12.8	555	1	INV1_KLUMA	P28999 kluyveromyc
7	359.5	12.6	532	1	INV2_YEAST	P00724 saccharomyc
8	354	12.4	477	1	CSCA_ECOLI	P40714 escherichia
9	348	12.2	581	1	INV1_SCHPO	O59852 schizosacch
10	330	11.6	476	1	RAFD_ECOLI	P16553 escherichia
11	320	11.2	533	1	INV1_DEBOC	P24133 debaryomyce
12	312	11.0	590	1	INV4_MAIZE	P49174 zea mays (m
13	276	9.7	512	1	INV4_ZYMMO	P35636 zymomonas m
14	274.5	9.6	636	1	INV4_LYCES	P29000 lycopersico
15	260.5	9.2	640	1	INV1_CAPAN	P93761 capsicum an
16	249	8.8	642	1	INV4_VICEA	Q43857 vicia faba
17	248	8.7	651	1	INV4_PHAVU	O24509 phaseolus v
18	245.5	8.7	465	1	SCRB_KLEPN	P27217 klebsiella
19	246.5	8.7	555	1	INV1_PEA	Q43089 pisum sativ
20	246	8.6	661	1	INV4_DAUCA	P80065 daucus caro
21	243	8.5	670	1	INV1_MAIZE	P49175 zea mays (m
22	242	8.5	592	1	INV1_DAUCA	P26792 daucus caro
23	242	8.5	649	1	INV4_PHAUU	P29001 phaseolus a
24	239.5	8.4	511	1	SCRB_ZYMMO	P22632 zymomonas m
25	238.5	8.4	432	1	BRAA_THEMEA	Q33833 thermotoga
26	237.5	8.3	494	1	SCRB_STAXY	Q05936 staphylococ
27	232.5	8.2	484	1	SCRB_VIBAL	P13394 vibrio algi
28	222	7.8	454	1	SCRB_STRMU	P13522 streptococ
29	214	7.5	466	1	SCRB_SALTY	P37075 salmonella
30	213	7.5	473	1	SCRB_LACIA	Q04937 lactococcus
31	206	7.2	501	1	SCRB_PEDPE	P43471 pediococcus
32	204.5	7.2	583	1	INV3_DAUCA	Q39693 daucus caro
33	197.5	6.9	480	1	SCRB_BACSU	P07819 bacillus su

34	187.5	6.6	592	1	INV2_DAUCA	Q39692 daucus caro
35	125	4.4	1120	1	STFR_ECOLI	P76072 escherichia
36	122	4.3	3491	1	ERV1_SACER	Q03131 saccharopol
37	117	4.1	451	1	Y483_MVCTU	Q11149 mycobacteri
38	116.5	4.1	984	1	EPA3_RAT	O08680 rattus norv
39	114.5	4.0	890	1	NTA2_PHAVU	P39866 phaseolus v
40	114	4.0	1132	1	VHSJ_LAMB	P03749 bacterioph
41	113.5	4.0	909	1	NIA_PETHY	P36859 petunia hyb
42	113	4.0	296	1	ALYS_BPDP1	O03979 bacterioph
43	112	3.9	916	1	NIA1_ORYSA	P16081 oryza sativ
44	111.5	3.9	752	1	HISA_PROAC	Q59634 propionibac
45	110.5	3.9	983	1	EPA3_HUMAN	P29320 homo sapien

ALIGNMENTS

RESULT 1
SACC_BACSU
ID SACC_BACSU STANDARD; PRT; 677 AA.
AC P05656;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Levanase precursor (SC 3.2.1.65) (2,6-beta-D-fructan
DE fructanhydrolase).
GN SACC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87286401; PubMed=3112519;
RA Martin I., Debarbouille M., Ferrari E., Klier A., Rapoport G.;
RT "Characterization of the levanase gene of Bacillus subtilis which
RT shows homology to yeast invertase.";
RL Mol. Gen. Genet. 208:177-184(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=88067786; PubMed=3120151;
RA Schoengenderfer K., Schwab H., Lafferty R.M.;
RT "Nucleotide sequence of a cloned 2.5 kb PstI-EcoRI Bacillus subtilis
RT DNA fragment coding for levanase.";
RL Nucleic Acids Res. 15:9606-9606(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97286548; PubMed=9141695;
RA Parro V., San Roman M., Galindo I., Purnelle B., Bolotin A.,
RA Sorokin A., Mellado R.P.;
RT "A 23911 bp region of the Bacillus subtilis genome comprising genes
RT located upstream and downstream of the lev operon.";
RL Microbiology 143:1321-1326(1997).
RN [4]
RP SEQUENCE OF 1-10 FROM N.A.
RX MEDLINE=90355183; PubMed=2117666;
RA Martin-Verstraete I., Debarbouille M., Klier A., Rapoport G.;
RT "Levanase operon of Bacillus subtilis includes a fructose-specific
RT phosphotransferase system regulating the expression of the operon.";
RL J. Mol. Biol. 214:657-671(1990).
CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 2,6-beta-D-
CC fructofuranosidic linkages in 2,6-beta-D-fructans (levans)
CC containing more than 3 fructose units.
CC -!- INDUCTION: IN CONTRAST WITH LEVANUCLEASE AND SUCLEASE, LEVANASE
CC SYNTHESIS IS NOT INDUCIBLE BY SUCROSE: NO INDUCER IS KNOWN FOR
CC LEVANASE.
CC -!- MISCELLANEOUS: LEVANASE CANNOT BE DETECTED IN THE WILD-TYPE
CC B.SUBTILIS BUT IS MOSTLY SECRETED INTO THE CULTURE MEDIUM BY SACL
CC MUTANTS, ESPECIALLY AT THE END OF THE EXPONENTIAL GROWTH PHASE.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X05649; CAA29137.1; ALT_INIT.
DR EMBL; Y00485; CAA68542.1; -
DR EMBL; X92868; CAA63465.1; -
DR EMBL; X56098; CAA39581.1; -
DR EMBL; Z99117; CAB44645.1; -
DR PIR; S06353; S06353.
DR PIR; A27286; A27286.
DR Subtilisin; BG10320; sacC.
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolyase; Glycosidase; Signal; Complete proteome.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 677 LEVANASE.
FT ACT_SITE 49 49 BY SIMILARITY.
FT CONFLICT 658 658 Q -> L (IN REF. 2).
SQ SEQUENCE 677 AA; 75951 MW; 80PD6B0A5EE7F525 CRC64;

Query Match 19.5%; Score 556; DB 1; Length 677;
Best Local Similarity 32.2%; Pred. No. 1e-32;
Matches 160; Conservative 74; Mismatches 217; Indels 46; Gaps 18;

QY 40 RAYVHTPPSGWCLDQRPVTHGAVQLYLLHSDONN--CPGGWDHASTTGDGVAFTHHGT 97

DB 35 RPOYHFTPEANMNDPNMGVYVYAGEVHLFYQHPYGLQWPMHGHAVSKDLTWHEH--L 92

QY 98 VMLRPDP--FPVNSGSAVGTANTAGFGAG---AVVALATQPTDGVKRYQEOVLYKSTDG 152

DB 93 PVALYPDEKTIIFSGSAVDKNTSFPQTKCKPLVAIYTDREG---HGVQIAVSNDK 149

QY 153 GFTFTALP-DPVIINTDGRAATTPAEIEAWEFRDPKIHMDTARGWCVIGRLRYAAFY 211

DB 150 GRWTWKYAGNPVIPN-----PGKKD---FRDPKVFWEKEKKVMVLAAGDRILLY 197

QY 212 TSNLRLDWTLRNDFPNHALGI-ECPLDFEITAD--DGRHVLVAASHDYGIGLPMPT 268

DB 198 TSNLRLDWTLRNDFPNHALGI-ECPLDFEITAD--DGRHVLVAASHDYGIGLPMPT 257

QY 269 YAYWTGTWDEQFQHADDLPQ--WLDGWDMYAAVTPSIDAPETKRLATAWNNNNKYAA 326

DB 258 MQYFVGDFDGTHEKNENPPNPKVLTWDYGRDFYAAVNSWDIPSTDRLWLGNWNNQY-A 316

QY 327 RDVPTDASDYGNGSIVRELRLARPGGWYTLSTPVAAL/TNYVTATTLPLDPTVDGSA 386

DB 317 NDVPTSP--WRSATSPRELKKAFTG-VRVVTQPVKLEITIRGTSKKWKNLTISPAS 372

QY 387 ---VLPNGRAYEI--ELDIAWDATNATNGISVGRSPDGTHTNIGKYGADLYVDGRPSDL 441

DB 373 HNYLAGSGDAYEINAEFFKVSAAEFKPKVRTGENQFTKVGVDYDRNKLFDVRSSEGN 432

QY 442 AGYSLA-PYSRAAAPIDPGARSVHLRLVDTQSVQVFNAGHGVLSQQVHFAGDGTGISL 500

DB 433 DTENPAFNTGKETAPLKPNGVKVLRIFVDRSSVEFVGNQGVITDIIILPDRSSKGL 492

QY 501 YDGGGAHFTGIVVREI 517

DB 493 YAAANG-----GVRVKSL 504

RESULT 2

FRUA_STRMU

ID FRUA_STRMU STANDARD; PRT; 1423 AA.

AC Q03174;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Fructan beta-fructosidase precursor (EC 3.2.1.80) (Exo-beta-D-
DE fructosidase) (Fructanase).
GN FRUA.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=93014178; PubMed=1398976;
RA Burne R.A., Penders J.E.C.;
RT "Characterization of the Streptococcus mutans GS-5 fruA gene encoding
RL exo-beta-D-fructosidase";
RL Infect. Immun. 60:4621-4632(1992).
CC -!- FUNCTION: THIS PROTEIN IS A FRUCTANASE ENZYME WHICH DEGRADES
CC LEVANS AND INULINS TO FRUCTOSE AND ALSO CLEAVES SUCROSE INTO
CC GLUCOSE AND FRUCTOSE AND CAN THEREFORE FUNCTION AS AN
CC EXTRACELLULAR INVERTASE.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal-non-reducing 2,1- and
CC 2,6-linked beta-D-fructofuranose residues in fructans.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan
CC by an amide bond (potential).
CC -!- INDUCTION: BY SUCROSE, FRUCTAN SUBSTRATES AND FRUCTOSE.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U78296; AAA26889.1; -
DR InterPro; IPR003343; Big_2.
DR InterPro; IPR001362; GH_32.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR Pfam; PF02368; Big_2; 1.
DR TIGRams; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; FALSE_NEG.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; FALSE_NEG.
KW Hydrolyase; Glycosidase; Cell wall; Peptidoglycan-anchor; Signal.
FT SIGNAL 1 39 POTENTIAL.
FT CHAIN 40 1391 FRUCTAN BETA-FRUCTOSIDASE.
FT PROPEP 1392 1423 REMOVED BY SORTASE (POTENTIAL).
FT ACT_SITE 458 458 BY SIMILARITY.
FT DOMAIN 867 871 INVOLVED IN BINDING OF SUGARS WITH
FT BETA-(2,6) LINKAGES OR BINDING OF
FT MOLECULAR WEIGHT FRUCTANS (BY
FT SIMILARITY).
FT SITE 1388 1392 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1391 1391 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 1423 AA; 158668 MW; 8E574715F4E72A8A CRC64;

Query Match 13.4%; Score 382.5; DB 1; Length 1423;
Best Local Similarity 28.2%; Pred. No. 8e-20;
Matches 148; Conservative 61; Mismatches 193; Indels 123; Gaps 27;

QY 40 RAYVHTPPSGWCLDQRPVTHGAVQLYLLHSDONNPGGWDHASTTGDGVAFTHHGT 98

DB 444 RDQYHYSVKDGMANDPNGLVYNGVHLHFQFYDDTKWGPMMHWAHATSTDLI-----HWKE 499

QY 99 MPLR--PDRP--VWGSAAVVGTAATAGF---GAGAVVALATQPTDGVKRYQEOVLYWSTD 151

DB 500 EPIAFYPDNSNGVWFSCVGVVDEHNSGLFKTAGGLVAITANGNG-----QRMELAYSED 555

QY 152 GGFTFTALPDPIVINTDGRAATTPAEIEAWE-----FRDPKTI-HWDTARGWCVV 201

DB 556 EGKTKQKY-DRIV-----ADWSNDPLQNDFRDPKVFHWN---NQFMV 595

QY 202 I--GRLYAAAFYTPSNLRLDWTLRNFDYPN-HALGGIECPDLFEITADGTRHVLAA 258

```
Db 596 LAGGRLR---YSSNNLKDWKVEST---YDPLHT-----ECPDMPIVANDGVVLKWLSSRG 646
QY 259 DAYGIGLPMYAYWGTGDEGFHADL---TPQMLDWCWDAVAWTW-----PS 305
Db 647 REYKVG---DFKQVDGKW---TFIADDAKDKDQVMNFKGDSYAAMTYVYVDFGTETRPT 700
QY 306 IDAPETKRLAIAMNNWKAARDVPTDASDGYNGQNSIVRELRLARQPCGWTLLSTPVA 365
Db 701 I-----PKTEVNMNTWEDYCNLVAADVTVGQDFNGFNLDLGLINE-NGQVILTQTPVK 755
QY 366 ALTNVVTATTL---PDRTVGSAVL---PWNGRAYEIELDIAWDATNVGISVGRSPDGPR 421
Db 756 AYDSLRYNTALHFKDVTDVDAANTLLKDFKDSYEVSHFRDEKTT-----802
QY 422 HFNICKYGADLYVDRGSPDSLACYSLAP-----YSRAAAPIDP-----GARS 462
Db 803 -----KVGFNLRVNGQATKVIYDLQETLTSIDRSQSGTILSAFAKVNQSHVTKNADGS 857
QY 463 VHLRLVDTQSVVEFVNAGHTVLSQOVHFAEGDGTISLYTDGCPA 507
Db 858 IDLHIYDRASVEFVSKNNTVAGANQIFPNPEAVGASIIVEGKA 902

RESULT 3
INVL_HANAN
ID INVL_HANAN STANDARD; PRT; 550 AA.
AC P40912;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Invertase precursor (EC 3.2.1.26) (Beta-fructofuranosidase 2)
DE (Saccharase).
GN INVL.
OS Hansenula anomala (Yeast) (Candida pelliculosa).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4927;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96163052; PubMed=8595669;
RA Perez J.A., Rodriguez J., Rodriguez L., Ruiz T.;
RT "Cloning and sequence analysis of the invertase gene INV 1 from the
RT yeast Pichia anomala.";
RL Curr. Genet. 29:234-240(1996).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X80640; CAA56684.1; -
DR PIR; S48812; S48812.
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 22
FT CHAIN 23 550
FT INVERTASE.
FT CARBOHYD 112 112
FT CARBOHYD 113 113
FT CARBOHYD 119 119
FT CARBOHYD 165 165
FT CARBOHYD 211 211
FT CARBOHYD 237 237
FT CARBOHYD 333 333
FT CARBOHYD 364 364
FT CARBOHYD 364 364
```

```
FT CARBOHYD 398 398
FT CARBOHYD 420 420
SQ SEQUENCE 550 AA; 63188 MW; 192BCD722455E121 CRC64;
Query Match 13.3%; Score 379; DB 1; Length 550;
Best Local Similarity 24.6%; Pred No. 4.5e-20;
Matches 139; Conservative 87; Mismatches 212; Indels 128; Gaps 27;
QY 39 LRAVYHMTPPSGMCLDQPPVTHGAY-----OLYYLHSDQNG-----PGWDHAST 86
Db 27 LRQIHLTPDQGMNDP-----NGMFYDRKDKLWHVYFOHNPDKKSIWATPVTWGHSTS 80
QY 87 TDGVAFTHTGTYM-PLRPDPFVWGSASVGTANTAGFGAGA-----VVALATQPTDGYR 139
Db 81 KULLTWDYHGNALPENDEDEGIFSGSVYDRNNTSGFFNDSTDPDEQRIVAITNNA-----136
QY 140 KYQEQLVYWSGTGGFTTALPDPIVINTDGRAATTPAEIEENAEWFRDPIKHWDTARGEW 199
Db 137 QLOTOEIAVSLDKGYSFYIKYDQNPVIN-----NSSQQRDPKVLWHDESNQWI 184
QY 200 CVIGRLR--YAAFYTSPNLRDWTLRNFDPYPNHALGGIECPDLFEITA-----DGTRHW 252
Db 185 MVAKTQEFKQVIYGSPLDKKDLKSNFTSNGYLFQYECPGFLKLPENLNDTVTSKW 244
QY 253 VLAASMDAYCIGLPM---TYAYWTGTWGDGEQPHADDLTLPQWLDWGDWYAAVTPSIDAP 309
Db 245 VLLLAIRP---GSLPGGSINEYFIGDFGTFPHDPDGGATRFMDIGKDFAFOSFDNTE-P 300
QY 310 ETKRLAIAWNNWKAARDVPTDASDGYNGQNSIVRELRLAR---OPGGM-VTLTSTPVA 365
Db 301 EDGALGLANWQY-ANTVPT---ENWRSSSLVRNYTLKYVDVNPENYGLTIQKPV- 355
QY 366 ALTNVVTATTL-----PDRTVGSAVLPMNGRAYE-----IELDIATWT 405
Db 356 ----YDKETRLNETLKTLETINEYEVNDLKDKSSFATDFNTERNATGVFEFLKF---409
QY 406 ATNVGISVGRSPDGRHTNIGY-----GADLYVDRGP-----438
Db 410 -TOTDLKMGYS---NMTQFGLYIHSQTVKGSOETLQLVDFDLSTTWYIDRTQHSFORN 465
QY 439 SDLAGYSLAPYSRAAAPIDPGARSVH-LRILVDTSQVEFVNAAGHTVLSQOVHFAEG--D 495
Db 466 SPVFTERTISTYVEKIDTTDQG--NVYTLGVVDNRNILELYFNDGSIAMNTFFFRGKIP 523
QY 496 TGISLYTDGGAHFT--GLIVREIGQ 519
Db 524 TSFEVVCDSEKSFITIDELSVRELAR 549

RESULT 4
INVL_YEAST
ID INVL_YEAST STANDARD; PRT; 532 AA.
AC P10594;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Invertase 1 precursor (EC 3.2.1.26) (Beta-fructofuranosidase 1)
DE (Saccharase).
GN SUC1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-73 FROM N.A.
RX MEDLINE=88216256; PubMed=2835632;
RA Hohmann S., Gozalbo D.;
RT "Structural analysis of the 5' regions of yeast SUC genes revealed
RT analogous palindromes in SUC, MAL and GAL.";
RL Mol. Gen. Genet. 211:446-454(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA Hohmann S.;
```

Submitted (NOV-1988) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
DR EMBL; X07570; CAA30457.1; -
DR PIR; S27372; S27372.
DR SGD; L0002143; SUC1.
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolyase; Glycosidase; Glycoprotein; Multigene family; Signal.
FT SIGNAL 1 19
FT CHAIN 20 532
FT ACT_SITE 42 42
FT CARBOHYD 23 23
FT CARBOHYD 64 64
FT CARBOHYD 111 111
FT CARBOHYD 112 112
FT CARBOHYD 118 118
FT CARBOHYD 165 165
FT CARBOHYD 275 275
FT CARBOHYD 356 356
FT CARBOHYD 369 369
FT CARBOHYD 384 384
FT CARBOHYD 398 398
FT CARBOHYD 512 512
FT CARBOHYD 532 532
SQ SEQUENCE 532 AA; 60570 MW; 3280EP260EA01606 CRC64;
Query Match 13.0%; Score 369.5; DB 1; Length 532;
Best Local Similarity 25.1%; Pred. No. 2.1e-19;
Matches 140; Conservative 84; Mismatches 221; Indels 113; Gaps 25;
QY 19 LALIFGAVPPAASGLRAVYHTPPSGWLCDPQR--PVTHGAYOLYVHSDQNN 76
DB 7 LFLLAGFAKISASWTNETSDRLVFTNKGMMNDPGLWDKAGKWHILFYQNPNDT 66
QY 77 G---PGWDHASTDGVATHTHG-----TVMPLRPDPVWMSGAVGTANTAGGAGAV-- 127
DB 67 VWGLPLFWGHATSDD---LTHWDEPVAIAPKRKDSGAYSGSMVIDYNNNTSGFFNDTIDP 123
QY 128 ---VALATQPTDGVKRYQEOYLYWSTDGGFTTAL-PDPVIVNTDGRAATTPAEIENAE 182
DB 124 RQRCVAIWYNTD---ESEQYISYSLDGGYTFTEYQKNPVL-----ANST 167
QY 183 WFRDPKHWDARGVWCVIGRLR--YAAFTYSPNLRDWTLRNFDYPNHALGG--IECP 238
DB 168 QFRDPKVFVYPSKKWIWTAANKSODYKIEIYSDLLKSKWLESAP--ANEGLGYQYEC 225
QY 239 DLFEITA--DQGRHVLWLAASMDAYGIGLPMTYA---YWTGTWGDGQFHADLTPOWLWD 293
DB 226 GLIEVPSEQDPKSHWVMSINP---GAPAGSGRNPQYFVSGFNGHFEAFEDNQSRVDF 282
QY 294 GWDYAAVTPSIDAPETRLAIAWNNKYYAARDVPTDASDGYNCQNSIVRELRLAROP 353
DB 283 GKDYALQTFNTDPTTYGALGIAWASNWEYSA-FYP---SNPWRSSMLSVLPFSLNT- 337
QY 354 GGVYLLSPVVAALTNVYATTTLPDRTVDGSAVLPPN-----GRAYEIEDIAW 403
DB 338 ---YQANPELNLKA-----EPILNISAGPWSRFATNTLTKANSNVNDLSNT 386
QY 404 DT-----ATNVGIVSGRS-----PDGTRHTNIGKYADLYVDRGFSDLA 442
DB 387 GTLEFELVAVVNTQTISKVSFADLSLWPKGLDEPPEYLRMGFEVSASSFELDGRGNKVK 446

QY 443 GYSLAPY--SRAAAPIDPGARSVHLRI-----LVDTQSVSEVFNAGHTVLSQQVHFAEGD 495
DB 447 FVKENPYFTNRMSVNNQPKSENDLSYKYVYGLLDQDQMLELYENDGDGVVSTNTYFTMTGN 506
QY 496 -----TCIS--LYTD 503
DB 507 ALGSVNNVTGVDNLFYID 524
RESULT 5
INVA_YEAST STANDARD; PRT; 532 AA.
AC P10596;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE Invertase 4 precursor (EC 3.2.1.26) (beta-fructofuranosidase 4)
DE (Saccharase).
GN SUC4.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-73 FROM N.A.
RX MEDLINE=88216256; PubMed=2835632;
RA Hohmann S.; Gozalbo D.;
RT "Structural analysis of the 5' regions of yeast SUC genes revealed
RT analogous palindromes in SUC, MAL and GAL";
RL Mol. Gen. Genet. 211:446-454(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA Hohmann S.;
RL Submitted (NOV-1988) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X07572; CAA30459.1; -
DR PIR; S27373; S27373.
DR SGD; L0002146; SUC4.
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolyase; Glycosidase; Glycoprotein; Multigene family; Signal.
FT SIGNAL 1 19
FT CHAIN 20 532
FT ACT_SITE 42 42
FT CARBOHYD 23 23
FT CARBOHYD 64 64
FT CARBOHYD 97 97
FT CARBOHYD 111 111
FT CARBOHYD 112 112
FT CARBOHYD 118 118
FT CARBOHYD 165 165
FT CARBOHYD 256 256
FT CARBOHYD 275 275
FT CARBOHYD 356 356
FT CARBOHYD 369 369
FT CARBOHYD 384 384
FT CARBOHYD 398 398
FT CARBOHYD 512 512
FT CARBOHYD 532 532
SQ SEQUENCE 532 AA; 60575 MW; A56EA2E3A49125EB CRC64;
Query Match 12.9%; Score 366.5; DB 1; Length 532;

Best Local Similarity 26.2%; Pred. No. 3.4e-19;		Matches 143; Conservative 80; Mismatches 231; Indels 91; Gaps 24;	
QY	21	LIFGAVPPAARASAPGSLRAYVHTPPSGWLCDDPOR--PVTHGAYQYLYLHSDONNG-	77
Db	9	LLAGFAAKISALMTWETSRLPLVHTPPKGMWMDPNGLWYDAKEGKWHLYFOYNPNDTW	68
QY	78	--PGGDHASTTGDGVAFTHG-----TVMLPRDPFPVWGSVAVGTANTAGGAGAV----	127
Db	69	GLPFWGHATSND--LTHWQDEPVAIAKRNDSGAYSGSMVIDHNTSEFFNDTVDPRQ	125
QY	128	--VALATQTDGVRKYQOYLWSTDGGFTTAL-PDPVIVMTDGRAATTPAEIENAEWF	184
Db	126	RCVAIWTYNTP--ESEEQYISLDDGGYTFTEYQKNPVL-----ANSTQF	169
QY	185	RPDKHMDTARGEMVCVIGRLR--YAAFTSPNLRDWTLLRRNFDVPHALGG--IECPDL	240
Db	170	RDPKVFWEPSQKWTMTAAKSQDYKIEIYSSDDLKSWKLESF--ANEGFLGQYECPL	227
QY	241	FEITA--DDGTRHWLAASMDAYGIGLPMYA---YWTGTWBGQFHADDLTPOWLDMGW	295
Db	228	IEVPTQDPSKSHWVFISINP---GAPAGGSFNQYFVGSFNGTHPEAYDNGSRVVDFCK	284
QY	296	DNYAAVTPSIDAPETKRLAIAMNWKYAADVPTDASDGYNGQNSIVRELRL-----A	350
Db	285	DIYALQTFNTDPTYGSGALGIAMSNWEYSA-FVPTNP---WRSSMSLVKRFSLNTEYQA	340
QY	351	RQPGGWYLLTLPVAAALTN-----YVATTTLPTDRTVDSAVLPWNGRAYEIELDAWTD	405
Db	341	NPETELINKAEPILNISAGPWLHFASNTL--TKANSFVLDLSGTGLEFELVYAVNT	399
QY	406	AFNVGISVG-----RSPDGRHTNIGKYADLYVDRGPSDLAGYSLAPY--SRAA	453
Db	400	TQSVKSVFSDLSLWFKGLEDEPEYLRMGFEASASSFFLDGRNSKVVKFKNPYFNIRMS	459
QY	454	APIDGARSVHURI-----LVTQSVFVFNAGHTVLSQQVHFAEGD-----TGIS-	499
Db	460	VNNQPFKSENDLSYKVKVGLLDQNLILEYFDNGDVSTNTYETMTGNALGSVMNTTGVDN	519
QY	500	-LYTD 503	
Db	520	LEYID 524	
RESULT 6			
ID	INUL_KLUMA	STANDARD;	PRT; 555 AA.
AC	P28999;		
DT	01-DEC-1992 (Rel. 24, Created)		
DT	01-DEC-1992 (Rel. 24, Last sequence update)		
DT	15-DEC-1998 (Rel. 37, Last annotation update)		
DE	Inulinase precursor (EC 3.2.1.7) (2,1-beta-D-fructanfructanohydrolase)		
DE	(Inulase).		
GN	INUL.		
OS	Kluyveromyces marxianus (Yeast) (Kluyveromyces fragilis).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Kluyveromyces.		
OX	NCBI_TaxID=4911;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE OF 24-56.		
RC	STRAIN=ATCC 12424;		
RC	MEDLINE=913372407; PubMed=1840529;		
RA	Laloux O., Cassart J.-P., van Beeumen J., Delcour J., Vandenhaute J.;		
RT	"Cloning and sequencing of the inulinase gene of Kluyveromyces		
RT	marxianus var. marxianus ATCC 12424."		
RL	FEMS Lett. 289:64-68(1991).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CBS 6556;		
RC	MEDLINE=94128347; PubMed=7764385;		
RA	Bergkamp R.J.M., Bootsman T.C., Toschka H.Y., Mooren A.T.A., Kox L.,		
RA	Verbakel J.M.A., Geerse R.H., Planta R.J.;		
RT	"Expression of an alpha-galactosidase gene under control of the		
homologous inulinase promoter in Kluyveromyces marxianus."			
RT	Appl. Microbiol. Biotechnol. 40:309-317(1993).		
RN	[3]		
RP	SEQUENCE OF 24-43.		
RA	MEDLINE=92304047; PubMed=2135869;		
RA	Rouwenhorst R.J., Hensing M., Verbakel J., Scheffers W.A.,		
RA	van Dijken J.P.;		
RT	"Structure and properties of the extracellular inulinase of		
RT	Kluyveromyces marxianus CBS 6556."		
RL	Appl. Environ. Microbiol. 56:3337-3345(1990).		
CC	FUNCTION: HAS BOTH INULASE AND INVERTASE ACTIVITY.		
CC	CATALYTIC ACTIVITY: Endohydrolysis of 2,1-beta-D-fructosidic		
CC	linkages in inulin.		
CC	SUBCELLULAR LOCATION: Secreted.		
CC	SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	EMBL; X57202; CAA40488.1; -		
DR	EMBL; X68479; CAA48500.1; -		
DR	PIR; S17502; S17502.		
DR	InterPro: IPR001362; GH_32.		
DR	Pfam: PF00251; Glyco_hydro_32; 1.		
DR	PROSITE: PS00609; GLYCOSYL_HYDROL_F32; 1.		
KW	Hydrolase; Glycosidase; Glycoprotein; signal.		
FT	SIGNAL 1 15 OR 16, OR 17 (POTENTIAL).		
FT	PROPEP 16 23 POTENTIAL.		
FT	CHAIN 24 555 INULINASE.		
FT	ACT_SITE 53 53 BY SIMILARITY.		
FT	CARBOHYD 32 32 N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 122 122 N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 123 123 N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 129 129 N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 176 176 N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 220 220 N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 226 226 N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 255 255 N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 290 290 N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 385 385 N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 393 393 N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 490 490 N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 549 549 N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	VARIANT 3 3 F -> L (IN STRAIN CBS 6556).		
FT	VARIANT 165 165 E -> Q (IN STRAIN CBS 6556).		
FT	VARIANT 168 168 T -> S (IN STRAIN CBS 6556).		
FT	VARIANT 251 251 V -> A (IN STRAIN CBS 6556).		
FT	VARIANT 252 252 D -> DS (IN STRAIN CBS 6556).		
FT	VARIANT 299 299 G -> D (IN STRAIN CBS 6556).		
FT	VARIANT 479 479 T -> N (IN STRAIN CBS 6556).		
FT	VARIANT 24 24 T -> S (IN REF. 3).		
FT	CONFLICT 43 43 H -> Y (IN REF. 3).		
FT	CONFLICT 43 43		
SQ	SEQUENCE 555 AA; 62213 MW; 723BAADDC3BF0907 CRC64;		
Query Match 12.8%; Score 365.5; DB 1; Length 555;			
Best Local Similarity 27.0%; Pred. No. 4.3e-19;			
Matches 132; Conservative 72; Mismatches 181; Indels 103; Gaps 24;			
QY	40	RAVYHMTTPSGWLCDDPQRPVTHGA-----YQYLYLHSDQN---NGPGWDHASTTD	88
Db	39	RPSVHFTPSHGWNNDP-----NGLWYDAKEEDWHLYYQYNPAATITGTPLYNGHAVSKD	92
QY	89	GVAFTTHGTVM-PLRDPFPVWGSVAVGTANTAGF-----GAGAVVALATQPTDGV	138
Db	93	LTSWTQYGLASLGSGDDAGAFSGSMVIDYNTSGFNSVDPQRAVAVVTLTKGFS---	149
QY	139	RKYQYLYWSTDGGFTTALPDVIVNTDGRAATTAEIENAEWFRDKIHWDTAR---	195


```
DR 63 PNDTVGTPLEWGHATSDDLTNWDOPAIAPKRNDSGAFSGSMVVDVNTSGFFNDITD 122
KW Hydrolyase; Glycosylase; Carbohydrate metabolism.
FT ACT_SITE 39 BY SIMILARITY.
SQ SEQUENCE 477 AA; 54363 MW; 4856622AEA380EBD CRC64;

Query Match 12.4%; Score 354; DB 1; Length 477;
Best Local Similarity 27.2%; Pred. No. 2.4e-18;
Matches 135; Conservative 67; Mismatches 206; Indels 88; Gaps 23;

QY 43 YHMTSPSGWLCDPQRPVTHGAYQLYLLHSDNN--GPGGDHASTTDCGVAATHCTVMP 100
Db 28 FLHAPAGMNDPGLNFWNDRYHAFYQHHPMSEHMGPMHGHATSDDDMIHWOHE--PTA 85
QY 101 LRPDEP-----VWGSAAVVGTAAGFGAGAVVALATOPTDGVRYKQEOYLWSDGGFT 155
Db 86 LAPGDENDKDCGFCSSGSAVDNGLVSLIYTHVWLDGAGNDDAIR--EVOCLATSDG--- 140
QY 156 FTALPDPIVNTDGRAATTPABEINAEWFRDKIHWDTARGEWVCV----IGRLRYAAY 211
Db 141 -----IHFEKQGVILTPP---EGIMHFRDPKV-WREADTWMVYVGAKPDPGNTGOILLY 189
QY 212 TSPNLRDWTLRNFDYPNHALGG---TICPDLEITADDGTRHWVLA--SMDAYGIGL 265
Db 190 RGSSUREWTFDRVL---AHADAGESYMWBCDPFFSL-----GQHYLMFSPQGMNAEG--- 239
QY 266 PMTYAYWT-----GTWDGEQFHADDLTPQWLDGMDWYAAVTWPSIDAPETKRLATA 317
Db 240 ---YSYRNRFGSVIPGWSGRLFAQSGHTELDNGHDFYAP---QSFVAKDGRRIYV 293
QY 318 WMNNKYAARDVPTDASGYNGQNSIVRELRLARPGGWYTLSTPVAALTYVATTTL 377
Db 294 WMDME---SPMPS-KREGWAGMTLAREL---SESNG--KLQRPVHEAESLROOHOQSI 344
QY 378 PDRTVDSGSAVLPNGRAYEIELDIW--DTATNVGIVSGRSPDGTHTNIGYKGLADLYD 435
Db 345 SPRTISNKYVLOENAGAVEIQLOWALKNSDAEHYGLQGA-----GMRLYID 391
QY 436 RGSPLAGYSLAPYSR---AAAPIDPGARSVHLRLVLTQSVFVFNAGHTVLSQQVHF 491
Db 392 NQSERLVLMRYYPHENLDGYRSIPL-PQGDMLALRIFDTSSVEVFINDGEAVMSSRIYP 450
QY 492 AEGDTGISLYTDGGPA 507
Db 451 QPEERELSLYASHGVA 466

RESULT 9
INVL_SCHPO
ID INVL_SCHPO STANDARD; PRT; 581 AA.
AC O59852; P78891;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Invertase precursor (EC 3.2.1.26) (Beta-fructofuranosidase)
DE (Saccharase).
GN INVL OR SPCC191.11.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98204792; PubMed=9535817;
RA Tanaka N., Ohuchi N., Mukai Y., Ozaka Y., Ohtani Y., Tabuchi M.,
RA Bhuyian M.S., Fukui H., Harashima S., Takegawa K.;
RT "Isolation and characterization of an invertase and its repressor
genes from Schizosaccharomyces pombe."
RL Biochem. Biophys. Res. Commun. 245:246-253(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
```

RESULT 8

```
CSCA_ECOLI
ID CSCA_ECOLI STANDARD; PRT; 477 AA.
AC P40714;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sucrose-6-phosphate hydrolase (EC 3.2.1.26) (Sucrase) (Invertase).
GN CSCA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=EC3132;
RA Bockmann J.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS ENZYME ENABLES THE BACTERIA TO METABOLIZE SUCROSE
AS SOLE CARBON SOURCE.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
fructofuranoside residues in beta-D-fructofuranosides.
CC -!- PATHWAY: Sucrose metabolism.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81461; CAA57219.1; -.
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
```

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds R., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltyjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Calbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of *Schizosaccharomyces pombe*."
RL Nature 415:871-880(2002).
RN [3]
RP SEQUENCE OF 88-581 FROM N.A.
RC STRAIN=PR745;
RX MEDLINE=98162722; PubMed=9501901;
RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT "Identification of open reading frames in *Schizosaccharomyces pombe*
RL cDNAs.";
RN DNA Res. 4:363-369(1997).
RN [4]
RP CARBOHYDRATES
RX MEDLINE=90253381; PubMed=2187435;
RA Moreno S., Sanchez Y., Rodriguez L.;
RT "Purification and characterization of the invertase from
RT *Schizosaccharomyces pombe*.";
RL Biochem. J. 267:697-702(1990).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -!- PTM: THE CARBOHYDRATE CONTENT OF THIS PROTEIN IS 67%. THIS IS
CC COMPOSED OF EQUI-MOLAR AMOUNTS OF MANNOSE AND GALACTOSE. THERE IS
CC ALSO A SMALL AMOUNT OF GLUCOSAMINE PRESENT.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011433; BAA25684.1; -;
DR EMBL; AL049644; CAB41057.1; -;
DR EMBL; D89242; BAA13903.1; -;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 22
FT CHAIN 23 581
FT ACT_SITE 97 97
FT BY_SIMILARITY -;
FT CARBOHYD 37 37
FT N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 40 40
FT N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 46 46
FT N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 57 57
FT N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 62 62
FT N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 79 79
FT N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 168 168
FT N-LINKED (GLCNAC. . .) (PROBABLE).

FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 446 446 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 569 569 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CONFLICT 189 189 MISSING (IN REF. 3).
FT CONFLICT 190 190 H -> Y (IN REF. 3).
FT CONFLICT 195 195 A -> P (IN REF. 3).
FT CONFLICT 198 198 Q -> L (IN REF. 3).
FT CONFLICT 205 205 D -> N (IN REF. 3).
FT CONFLICT 243 243 I -> M (IN REF. 3).
FT CONFLICT 246 246 M -> L (IN REF. 3).
FT CONFLICT 256 256 S -> P (IN REF. 3).
SQ SEQUENCE 581 AA; 64407 MW; 70206A6CD1F27BC4 CRC64;
Query Match 12.2%; Score 348; DB 1; Length 581;
Best Local Similarity 24.9%; Pred. No. 8.3e-18;
Matches 132; Conservative 82; Mismatches 226; Indels 90; Gaps 22;
QY 28 PPAARSAFGSL-----RAYVHTPPSGWLCDDQRPVTHCAVQLYY 69
DB 53 PPEVNTAPNGTCLGNYNEPLSGYINATDRPKIHFTPPSGFMNDPGLYTGCVYHMF 112
QY 70 LHSDDQ--NNGPGGWDHASTDGTGVAF-----THHGTVMPRLRDPFVWMSGVAVGTA 117
DB 113 QKSPKLTAGEVHNGHTVSKDLHWNYPITAIYDPDEHENGVLSL-----PFSSAVVDVH 167
QY 118 NTAGFCAG-----AVVALATQPTDGVRYQEQYLXWSTDGGFTTALPDPIVNTDGRA 171
DB 168 NSSGLFSNDTIPERIVLIVTDHWTGVA--ERQAIATVTDGGYTFK-----KY 213
QY 172 ATPPAEIAENAEWRDPKIHWDTARGWCVGVIGRLY--AAFYSPNLRDWTLRNFDYPN 229
DB 214 SGNPVLIDINSLOFRPKVWDFDANRWYIVAMSONYAFYSYDLIHTELVSFSTG 273
QY 230 HALGGTECPDLFEITADGGTRH--WLAASMDAYGIGLPM---TYAYWTGWDGEGQHADD 285
DB 274 YGLQVECEGMARVPVEGDEYKWLFIINP---CAPLGSSVQVYFVGDMGTNEFVDD 330
QY 286 LTPQWLDWGDWYAAVTPSIDAPETKRLAIAMNNKYAARDVPTDASDGYNGQNSIVR 345
DB 331 GQTRFVDLGKDFVASALYHSSA--NADVIGVGWASNNQY--TNOAPTQV---FRSAMTVAR 385
QY 346 ELRLARQPGGWYT---LLSTP--VAALTN--VYTATTLTLDRTVDGSALVLPNGRAYE- 396
DB 386 KFTLRDVPQNPMTNLTSLQTLPLNLSLLRDETFLTAPVINSSSSLSGSPITLPSNTAFEF 445
QY 397 -IELDIAMDTATNNGISVGR-----SPDGR-----HTNIGKYGADLYVDGRGPSDIAGY-SL 446
DB 446 NVTLSINYEGTGTGCLGRIIDSDPYRLQSLISVDVDFAASTLVINRAKAGMGWNSL 505
QY 447 APYSRAAAPIDPGARSVHLRLVDTOSVEFVNAGHTVLSQQVHFAEGDT 496
DB 506 FTSPFANDIYIG--NVTLYGIVDGLLELYVNGEKTNTDFFLQGT 553
RESULT 10
RAFD_ECOLI
ID RAFD_ECOLI STANDARD; PRT; 476 AA.
AC P16553;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE Raffinose invertase (EC 3.2.1.26) (Invertase).
GN RAFD.
OS Escherichia coli.
OG Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.

Db 173 -SSNQPRDPKVFHRRPKSMHGCSE-----IARVKIQIF-GSANLKNWVLSNFSFGYYG 226
QY 230 HALGGIECPDLFEITADDGTR-HWVLAASMDAYGIGLPM---TYAYWTGTWDEQFHADD 285
Db 227 NOYG---MSRLIEVPIENSCKSWFLAIPN---GSPLGSSINQYFVGFDFGQFVPDD 280
QY 286 LTPQWLDWGWNTAAVTPSIDAPETKRLAIAMNNWKKYAADVPDASDGYNGQNSIVR 345
Db 281 SOTRFVDIGRDFYAFOTFSEV---EHGVLGLAWASNQYADQ-VPTNP---WRSSTSLAR 333
QY 346 ELRLARQPGGWYLLSTPVA---LTPNYVTTATTLDPDRTVGGSAVLP---WNGRAY 395
Db 334 NYTLR-----YVQMLKUTANIDKSVLPDSINVVDKLLKKNKLTNNKPIKTFNKGSTG 387
QY 396 EIEDIAWDTATNVGISVGR-----SPDGRTRHNIQIGYAD-----LYVDRG----- 437
Db 388 LFDENITF-KVLNUNVSPCKTHFDILINSQELNSSVDISKIGFSSQSLSFYIDRHIPNVE 446
QY 438 -----PSDLACYSLAPYSRAAAPDPGARSVHLRLVDVTSQVEFVNAGHTVLSQQVH 490
Db 447 FPRQFFTDKLAAY-LEPLD-----YDQDLRVFSLYGIVDKNIIELYFNDGTVMANTTF 500
QY 491 FAEG-----DTGISLYTDGGAHFTGIVWREIGQ 519
Db 501 MGEKYPHDIQIVTDEEPLFELESVIIRLKN 533
RESULT 12
INVA_MAIZE
ID INVA_MAIZE STANDARD; PRT; 590 AA.
AC P49174;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Beta-fructofuranosidase, cell wall isozyme precursor (EC 3.2.1.26)
DE (Sucrose-6-phosphate hydrolase) (Invertase).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Black Mexican Sweet;
RA Shanker S.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -! CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
fructofuranoside residues in beta-D-fructofuranosides.
CC -! SUBCELLULAR LOCATION: Cell wall.
CC -! INDUCTION: By wounding and bacterial infection.
CC -! SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U17695; AAA64487.1; -
DR MaltedB; I13032; -
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; FALSE NEG.
KW Hydrolase; Glycosidase; Glycoprotein; Cell wall; Signal.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 590 BETA-FRUCTOFURANOSIDASE, CELL WALL
ISOZYME.
FT ACT_SITE 68 68 BY SIMILARITY.
FT CARBOHYD 190 190 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 590 AA; 65198 MW; 296455E61E99B892 CRC64;

Query Match 11.0%; Score 312; DB 1; Length 590;
Best Local Similarity 25.3%; Pred. NO. 3.3e-15;
Matches 150; Conservative 81; Mismatches 211; Indels 150; Gaps 34;
QY 19 LALIFGA-----VPPAARASAPGS-----LRAYVHTPPSGWLCDDPQRPVTHGAYQ 66
Db 21 LALRLAGASHVHRSLEAEAPSPASIVSPLRLRTGYHFQPPMNNWINDPNAPLYYKGYH 80
QY 67 LYVLHSDONNGPG--WDHASTTGDVAFTH-HGTVMPLRPD--FPVWMSGSAVVGCTANTAG 121
Db 81 LFTQYKPGKAVGNIVVAHSVSRDLINWVALEPAIYPSISDKYGCWSGATI-----LE 135
QY 122 FGAGAVVALATQDGTGVRYQEQLYXWSTDGGFTTALP-----DPVIVNTD-----GRAA 172
Db 136 DGTTPAILYTGIDRAD--INVQVQL-----ALPKDASDPLLRWEKPEEYNPVA 182
QY 173 TTPAEIENAEWFRDPKIHWDTARGEVVCVIGRLR-----YAAFTYSPMLRDMTURLRNDYP 228
Db 183 TPAAGGINATQFRDPTTAMRHA-GHWRMLVSGVRGARGMALVYRSRDRFRKWKAK---HP 238
QY 229 NH--ALGGI-ECPLFEI-----TADDCGTRHWVLAASMDAYGIGLPMYVAYWT-G 274
Db 239 LHSAAITGMWECDFPVPVSGPGLQAGLDTSAPGK-YVLKSSLDL-----TRYDIYTTG 291
QY 275 TWDG--EQFHADDLTPQW-----LDWGWYAAVTPSIDAPETKRLAIAMNNWKKYAAAR 327
Db 292 SYDGGKDRYYPDPAGDYHRRRYDYG-NYYASKTF--YDPVERRRVLLGWANE---SD 344
QY 328 DVPTDASDNGYNGNSIVRELRLARQGGWYLLSTPVAALTNVYVATTTLPDRTVDGSAV 387
Db 345 SVPDKAKAGWAGIHAIPRKIWL--DPTG-KQLLQWPIHEVEKLRGKAV-----SVDKLV 396
QY 388 LPWNG-----RAYETELDIAMDTATNVGTSV-----GRSPDGTHTNIGYKADLYV 434
Db 397 KPCHFEVGTIATYQADVEVSFELEAGTSLEKAEAFDPAYDDDAQK:CGYKGDARG 456
QY 435 DRGP-----SDLAGYSLAP---YSRAAA---P 455
Db 457 GVPFGGLWLASADLQERTAVFVRFDGHHGKPKVLMCTDPTKSSLSPLDYKPTFAFVD 516
QY 456 IDPGARSVHLRLVDVTSQVEFVNAGHTVLSQQVH--FAEG-DTGISLYTDG 504
Db 517 ADISSGKITLRSLIDRSVVSFGAGGKTCILSRVYPSIAVGKDAHLVYFNNG 568
RESULT 13
INVA_ZYMO
ID INVA_ZYMO STANDARD; PRT; 512 AA.
AC P35636;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Sucrose-6-phosphate hydrolase E1 (EC 3.2.1.26) (Sucrase E1)
DE (Invertase E1).
GN INVA.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=Z6C;
RX MEDLINE=91291345; PubMed=1368686;
RX Yanase H., Fukushi H., Ueda N., Maeda Y., Toyoda A., Tomomura K.;
RT "Cloning, sequencing, and characterization of the intracellular
invertase gene from Zymomonas mobilis";
RL Agric. Biol. Chem. 55:1383-1390(1991).
CC -! CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
fructofuranoside residues in beta-D-fructofuranosides.
CC -! SUBCELLULAR LOCATION: Cytoplasmic.
CC -! SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----

[1] SEQUENCE FROM N. A.
RN RC STRAIN=cv. UC82B, and wild type;
RP STRAIN=cv. UC82B; PubMed=8095164;
RX MEDLINE=93184207; PubMed=8095164;
RA Elliott K.J., Butler W.O., Dickinso
RZ Fitzmaurice L., Mirkov T.E.;
RT "Isolation and characterization of
RN "Isolation and characterization of fruit vacuolar invertase genes

RI FROM TWO tomato species and temporal differences in mRNA levels
 RT during fruit ripening.";
 RL Plant Mol. Biol. 21:515-524 (1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Klann E.M., Yelle S., Bennett A.B.;
 RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-553 FROM N.A.

RX MEDLINE=93283182; PubMed=1304771;
 RA Ohyama A., Hirai M., Nishimura S.;
 RT "A novel cDNA clone for acid invertase in tomato fruit.";
 RC

SEQUENCE OF 474-636 FROM N.A.
STRAIN=cv. Ailsa Craig; TISSUE=Fruit;
[4]
Jpn. J. Genet. 67:431-432(1992).

Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-fructofuranoside residues in beta-D-fructofuranosides.

CC	-	FAHRAW: FIRST ENZYME INVOLVED IN THE METABOLISM OF INCOMING
CC	-	SUCROSE.
CC	-	SUBCELLULAR LOCATION: Vacuolar.
CC	-	SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES
CC	-	SIMILARITY:

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

CC BY-NC-SA. All rights reserved. This document is not for commercial use and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>) or send an email to license@isb-sib.ch.

CC	EMBL; Z12025; CAA78060.1; -
DR	EMBL; Z12026; CAA78061.1; -
DR	EMBL; Z12027; CAA78062.1; -
DR	EMBL; Z12027; CAA78062.1; -

DR	EMBL	Z12029	CAA76003.1	-
DR	EMBL	M81081	AAA34132.1	-
DR	EMBL	D11350	BAA01954.1	-
DR	EMBL	X77264	CAA54480.1	-

DR S28511.
DR S28513.
PIR S28514.
PIR S28515.
PIR S31157.
PIR S31158.
PIR S31159.

DR F1K, S31137; GH_32
DR InterPro; IPR001362; GH_32
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
DR

NW	hydrolase, glycosidase,
FT	glycoprotein; zymogen, signal.
FT	POTENTIAL.
FT	SIGNAL
FT	1
FT	47
FT	92
FT	PROPEP
FT	93
FT	CHAIN
FT	636
FT	ACID BETA-FRUCTOFURANOSIDASE.

FT	16	N-LINKED (GLCNAC. .)	(POTENTIAL).
CARBOHYD	144	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	211	N-LINKED (GLCNAC. .)	(POTENTIAL).
CARBOHYD	276	N-LINKED (GLCNAC. .)	(POTENTIAL).

CARBONITE 05
 FT CONFLICT 542
 SQ SEQUENCE 636 AA; 70097 MW; 74C46246FA93974B CRC64;
 SSEEAPGVKKVY -> PAFSLSTINUSL (IN REF. 3).
 DIBADU (DECA-17) (UNCLASSIFIED)

Query match 3.08; Score 274.5; Length 630;
 Local Similarity 24.1%; Pred. No. 1.8e-12;
 Matches 137; Conservative 76; Mismatches 236; Indels 119; Gaps 28;

QY 2 1 FALSERNAV LQSGRGORLALLEGGVFFRANSGRFGS LNAVTHMIFFGWZCCEQKPV1 00

```
Db 65 SPAPPSRGVSGVSDKTRFDVAGASHVSVAWSNMLSWORTAYHFQPKQNMNDPGLY 124
QY 61 THGAYOLYLHSDNN--GPGGDHASTTGDGVAFTHTGVTMP--LRPD-----FPVWSGS 111
Db 125 HKGWYHLFYQNPDSAINWITGHAVSKDLI-----HWLYLPFAMVPDQWYDINGVWTGS 180
QY 112 AVVGTANTAGFAGAGAVVALATOPTDGVKRYQEQYLXWSTDGFTTALPDVPI--VNTD 168
Db 181 ATT-----LPDGOIMLYTGTDDY--VQVONLAYPAN-----LSDPLLDWVKFK 224
QY 169 GRAATTPAEIAENAEWRDPKIIH--DTARGEWCV-----IGRLRYAAFTYSPNLRDWTLR 223
Db 225 GNPVLVPPPGIGVKDRDPTTAWTGPQNGQWLLTIGSKICKTGVALVYETSFTSKL-- 282
QY 224 NFDPYPHALGGI---ECPDLFEITA-----DDGTRHWVLAASM-----DAYGTG 264
Db 283 -LDGVLHAYVPGTGMWECVDFYFVSTKTKTNGLDTSYNGPGVKH--VLKASDDNNKQDHYAIG 340
QY 265 LPMTYAYWTGTWDEQFHADDLTPQWLDGWMDWYAAVTPSIDAPETKRLALAMNNWKY 324
Db 341 ---TYDLGRKNKTPDNPDLDCGIGLRLDYG--KYASKTF--YDPKKERRVLMGICETDS 394
QY 325 AARDVPTDASDYGNGNSIVRELRLARQPGWYTLTSTPVAALTNVVTATTLTPDRTVD-- 383
Db 395 ESADL-----OKGWSVQSIPRTVLYDKKTG--THLLQWPEVETESLURVGDPVT--KQVDL 446
QY 384 -GSAPLPWNGRAYETEIDAMDT-----ATNVGISVGRSPDGTTRHTNIGKYG-- 429
Db 447 QPGSIELLRVDSNAELDIASFEVDKVALOGIIEADHVGFSCTSGAASRGILGPGVI 506
QY 430 --AD-----LYVDRG-----PSDLAGYSLAPYSRAAIPDPGARS 462
Db 507 VIADQTLSELTPVYFISKADGAEHFCADQTRSRSEAPGVGKQVYGSVPVLDCGKHS 566
QY 463 VHLRLILVDTSQVEVFVNAGHTVLSQOVH 490
Db 567 --MRLLVDSHIVESFAQGGRVTITSRIY 592
```

RESULT 15

```
INVL_CAPAN STANDARD; PRT; 640 AA.
AC P93761.1
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Acid beta-fructofuranosidase AIV-18 (EC 3.2.1.26) (Acid sucrose-6-
DE phosphate hydrolase) (Acid invertase).
OS Capsicum annuum (bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fruit;
RA Choi D.; Lee K.-W.; Kim S.;
RT "Isolation and characterization of acid invertase cDNA clone in Hot
RT pepper (Capsicum annuum L.) fruits.";
RL J. Plant Biol. 40:298-303(1997).
CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -I- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL; U87849; AAB48484.1; -.
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Transmembrane; Glycoprotein.
FT TRANSMEM 34 54
FT ACT_SITE 126 126
FT CARBOHYD 152 152
FT CARBOHYD 219 219
FT CARBOHYD 491 491
FT CARBOHYD 612 612
SQ SEQUENCE 640 AA; 70620 MW; D3C628B7A7E6870B CRC64;
Query Match 9.2%; Score 260.5; DB 1; Length 640;
Best Local Similarity 23.8%; Pred. No. 1.9e-11;
Matches 138; Conservative 77; Mismatches 221; Indels 143; Gaps 32;
QY 1 MTPAISRRVAVLQAGAGALALIFGGAVPPAARASAP--GSLRAYVHMTPPSGWCLDQRPV 59
Db 72 LTPATPSRGVSQGVSEKTFKDSVTSQSVTSNAMLNMQRTAYHFQPKQNMNDPGL 131
QY 60 TTHGAYOLYLHSDNN--GPGGDHASTTGDGVAFTHTGVTMP--LRPD-----FPVWSG 110
Db 132 YHKGWYHLFYQNPDSAINWITGHAVSKDLI-----HWLYLPFAMVPDQWYDINGVWTG 187
QY 111 SAVVGTANTAGFAGAGAVVALATOPTDGVKRYQEQYLXWSTDGFTTALPDVPI--VNT 167
Db 188 SATI-----LPDGLIMLYTGTDDY--VQVONLAYPAN-----LSDPLLDWVKY 231
QY 168 DGRAATTPEIAENAEWRDPKIIH--DTARGEWCV-----IGRLRYAAFTYSPN--LRDWT 220
Db 232 QGNPVLVPPPGIGVKDRDPTTAWTGPQNGQWLLTIGSKICKTGVALVYETSFTSKL 291
QY 221 LRNFDPYPHALGGI---ECPDLFEITA-----DDGTRHWVLAASMDAYGILGP 266
Db 292 L-----HAPGTGMWECVDFYFVSTLDAAGLTSYNGPGVKH--VLKASLDD-----N 337
QY 267 MTYAYWTGTW--GEQFHADDLTPQW-----LDWGDWYAAVTPSIDAPETKRLAIA 317
Db 338 KODHYVIGTYDVPKKNFSPDN--PDLDCGIGLRLDYG--RYASKTF--YDPKKQRRVLWG 392
QY 318 WNNWKYAADVPTDASDYGNGNSIVRELRLARQPGWYTLTSTPVAALTNVVTATTL 377
Db 393 WIGETDSESADL-----OKGWSVQSIPRTVLFDKKTG--THLLQWPEVETESLURSGDPKV 446
QY 378 PDR-----VDGSAVLPWNGRAYETEIDAMDTAT-----NVGISVGRSPDGT 420
Db 447 KEVNLOPGSIELLHVDSSA-----QFDEASFEDRVTLLEGIIADYGYNCSTSGGAA 499
QY 421 RHTNIGKYG---AD-----LYVDRG-----PSDLAGYSLAPYSR 451
Db 500 SRGILGPGFVGVVIADQTLSELTPVYFISKADGAEHFCADQTRSRSEAPGVGKQVY 558
QY 452 AAAPIDPGARSVHLRLILVDTSQVEVFVNAGHTVLSQOVH 490
Db 559 SSVFVLDCGKHK--RMRLLVDSHIVESFAQGGRVTITSRIY 596
```

Search completed: May 6, 2003, 12:32:39

Job time : 14.5202 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: May 6, 2003, 12:31:02 ; Search time 19.5375 seconds
(without alignments)
784.611 Million cell updates/sec

Title: US-09-868-328b-1
Perfect score: 2845
Sequence: 1 MTPAIRRAVLQAGAGALA.....TDGGPAHFTGIVVREIQAI 521

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	499.5	17.6	943	4	US-09-397-885-5
2	488.5	16.5	1277	4	US-09-397-885-3
3	461	16.2	923	4	US-09-397-885-1
4	348	12.2	581	4	US-09-331-581-2
5	333.5	11.7	337	4	US-09-331-581-24
6	291	10.2	332	4	US-09-331-581-23
7	275	9.7	635	1	US-08-245-809-1
8	274.5	9.6	636	1	US-08-296-624-2
9	235.5	8.3	492	4	US-09-134-001C-3895
10	205.5	7.2	630	3	US-08-860-091A-2
11	205.5	7.2	630	4	US-09-245-323A-6
12	189	6.6	615	3	US-08-860-091A-4
13	189	6.6	615	4	US-09-245-323A-8
14	169.5	6.0	626	4	US-09-019-385-2
15	147.5	5.2	565	4	US-09-142-623-11
16	141.5	5.0	635	4	US-09-142-623-1
17	140.5	4.9	574	4	US-09-142-623-13
18	122	4.3	3491	2	US-07-642-734C-2
19	122	4.3	3491	3	US-08-439-009A-2
20	119.5	4.2	3739	3	US-09-320-878-2
21	115	4.0	3739	4	US-09-105-537-33
22	115	4.0	11877	4	US-09-105-537-6
23	110.5	3.9	983	1	US-08-167-919A-10
24	110.5	3.9	983	2	US-08-449-645A-21
25	110.5	3.9	983	2	US-08-702-367A-21
26	110.5	3.9	983	3	US-08-715-106-10
27	110.5	3.9	983	5	PCT-US95-04681-21

28	106	3.7	3519	4	US-09-428-517-4	Sequence 4, Appli
29	106	3.7	4551	3	US-09-320-878-1	Sequence 1, Appli
30	106	3.7	4613	4	US-09-105-537-31	Sequence 31, Appl
31	105.5	3.7	811	1	US-08-480-604A-7	Sequence 7, Appli
32	105.5	3.7	811	2	US-08-405-496A-7	Sequence 7, Appli
33	105.5	3.7	811	4	US-08-915-136-7	Sequence 7, Appli
34	105.5	3.7	811	4	US-08-957-310-7	Sequence 7, Appli
35	105.5	3.7	812	1	US-08-480-604A-29	Sequence 29, Appl
36	105.5	3.7	812	4	US-08-915-136-29	Sequence 29, Appl
37	105.5	3.7	2710	1	US-08-480-604A-6	Sequence 6, Appli
38	105.5	3.7	2710	2	US-08-405-496A-6	Sequence 6, Appli
39	105.5	3.7	2710	4	US-08-915-136-6	Sequence 6, Appli
40	105.5	3.7	2710	4	US-08-957-310-6	Sequence 6, Appli
41	104.5	3.7	964	4	US-08-484-791-2	Sequence 2, Appli
42	104	3.7	655	4	US-09-311-626B-14	Sequence 14, Appl
43	104	3.7	3033	1	US-07-925-695-5	Sequence 5, Appli
44	103	3.6	936	5	PCT-US94-05905-22	Sequence 22, Appl
45	102.5	3.6	614	1	US-08-262-338A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-397-885-5
; Sequence 5, Application US/09397885
; Patent No. 6323007
; GENERAL INFORMATION:
; APPLICANT: Moller, Soren
; APPLICANT: Johansen, Charlotte
; APPLICANT: Schafer, Thomas
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Hoeck, Lisbeth Hedegaard
; TITLE OF INVENTION: A 2,6-b-D-Fructan Hydrolase Enzyme And
; TITLE OF INVENTION: Process For Using The Enzyme
; FILE REFERENCE: 5540.200-US
; CURRENT APPLICATION NUMBER: US/09/397,885
; CURRENT FILING DATE: 1999-09-17
; EARLIER APPLICATION NUMBER: PA 1998 01173
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PA 1998 01623
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 60/101,615
; EARLIER FILING DATE: 1998-09-24
; EARLIER APPLICATION NUMBER: 60/111,675
; EARLIER FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Paenibacillus macerans
US-09-397-885-5

Query Match 17.6%; Score 499.5; DB 4; Length 943;
Best Local Similarity 30.8%; Pred. No. 1e-36;
Matches 157; Conservative 67; Mismatches 223; Indels 63; Gaps 19;
QY 40 RAVYHMTPPSGMCLDPQRPVTHGAYQLYLLHSDNNGPGWDHASTDDGVAFTHHGVIM 99
DB 383 RPDYHSPARGSASDPNGILVYGEYHLF--HQD---GGTWAHAVSTDLVHWKRLPIAL 436
QY 100 PLRPFDPVWSSGAVGTANTAGF----CAGAVVALATOPTDGVKRYQOYLWSTDDGF 154
DB 437 PMNDLGHVWSSGAVSADLNHNASGLFADSGKGLIAYTTSYNDPRPNCNQRIGLAYSKDRGR 496
QY 155 TPT-ALPDPIVINTDGRAATTPAELENAEW-FRDPCKIHWDTARGEWVCVIGRLRYAAFT 212
DB 497 TWYEAAREPIVIENTGKQDDP-----GGWDFRDKPVVDEEHNKRWVWVSGDHIREF 551
QY 213 SPNLRDWTLLRRNFDYPNHALGGI-ECPLDFEITADD-GTRHWVLAASMDAYIGILPMTYA 270
DB 552 STNLIIDWTLTDSFGYGVYRGVWECPLDLQLAVDDTGCKWVLMISTGAN-----PNTQG 607


```
Db 561 STNLLDWTLDNMGVDYVRGWCPCDLPV-DGTSQKKVAMISTGAN-----PKTG 615
Qy 269 ---YAWTGT--TWDEQFADDLTPOWL--DWGWDYAAVTPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGONSIVR 345
Db 616 GSDAEYFFGHLTADG-KFVNDNPACKVLTDFGKEFYASMSFANM--PDHRTVMMAWMTN 672
Qy 322 WKYAADVPTDASDGYNGONSIVRELRLARQPGWYTLSTPVAALTNVYVATTPLPDR 381
Db 673 WDYPEA-FPT---SNMKGELTIPREVSIVTTEG-IRMVQSPKEL-----ESLRKPLYS 722
Qy 382 VDSAVLPWNGR-----AYEIELDI---AWDTATWVGISVGRSPDGTTRHNTICKYGA 430
Db 723 ASNKSVSPSSGNLLKGIISGAYEIAEIEIPETSVTEFGFIREG--ANOKTVVGYKAS 780
Qy 431 D---LVVDRGPDLAGYSLAPYSRAAAPIDPGARSHLRILVDTQSVFVFNAGHTVLSQQ 488
Db 781 DSRMEVDRTASGETDPSNLSFKKHEAPTQOMENNRKMRILVDSESSVEAFNGDKGVVPSDV 840
Qy 489 VHFAGDGTISLYTDGGAHFTGIVVREI 517
Db 841 IPPDPASRAMSFYVKGNNVVVSLKVHQL 869

RESULT 4
US-09-331-581-2
; Sequence 2, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
; APPLICANT: TOHDA, Hideki
; APPLICANT: KUMAGAI, Hiromichi
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
; FILE REFERENCE: 0059-1142-0PCT
; CURRENT APPLICATION NUMBER: US/09/331.581
; EARLIER FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: PCT/JP98/04929
; EARLIER FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: JP 9-314608
; EARLIER FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 581
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-09-331-581-2

Query Match 12.2%; Score 348; DB 4; Length 581;
Best Local Similarity 24.9%; Pred. No. 3e-23;
Matches 132; Conservative 82; Mismatches 226; Indels 90; Gaps 22;

Qy 28 PPAARASAPGSL-----RAVYHMTPPSGWLCDDPQRPVTHGAYQIY 69
Db 53 PPVNTTAPNCTCLGNAYNEFLPSGYNATDRPKIHFTSPSGFMNDPGLVVTGGYHMF 112
Qy 70 LHSDD--NNGPGGWDAHSTDGAF-----THGTVMLRPLRDPVNSGSAVVGTA 117
Db 113 QYSPKTLTAGEVHGHGTVSKDLIHENYPIAIPDEHENGVLUL-----PFGSAVDVH 167
Qy 118 NTAGFAG-----AVVALQTPTDGRVKYQBYLYWSTDGGFTFTALPDPIVNTDGRA 171
Db 168 NSSGLFSNDTIPERIVLIYTDHWTGVA--ERQATAYTDDGGYTPK-----KY 213
Qy 172 ATTPAEIENAEWFRDPKTHWDTARGEWCVICRLRY--AAFYTSNLRDWTLRNFDYPN 229
Db 214 SGNPVLIDNSLQRPKPVINDFANDRWVMIAMSONYGIATFVSSYDLIHWTELSVFSTG 273
Qy 230 HALGGIECPDLFEITADGTRH--WVLAASMDAYGIGLPM---TYAYWTGTWDGEOFHADD 285
Db 274 YLGQYECPGMARPEVGTDEYKWLFLFISINP---GAPLGGSVQYFYVGDWNGTNFVPPD 330
```

```
Qy 286 LTPQWLDMGWDWYAAVTPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGONSIVR 345
Db 331 GQTRFVLDGKDFYASALYHSSSA-NADVTIGVGWASNQY--TQAPTQV---FRSAMTVAR 385
Qy 346 EURLARQPGGWT-----LLSTP--VAALTN--YVTTATTLPORTVDGSAVLDPWNGRAYE- 396
Db 386 KFTLRDVPQNPMTNLTSLIQTPLNVSLRLDETFLTPAPVINSSSSLSGSPITLPSNTAFEF 445
Qy 397 -TELDIANDTATNVGISVGR---SPDGR---HTNIGKYGADLYVDVRGPDLAGY-SL 446
Db 446 NVTLSINTEGCTGYCLGRIIIDSDDPYRLOSISVDVDFAASTLVLINRAKQMGWNSL 505
Qy 447 APYSRAAAPIDPGARSHLRILVDTQSVFVFNAGHTVLSQQVHFAGDT 496
Db 506 FTFSPANDIYIG--NVTLVGIVDNGLLLEYVNGEKTVTNDFFFLOGAT 553

RESULT 5
US-09-331-581-24
; Sequence 24, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
; APPLICANT: TOHDA, Hideki
; APPLICANT: HAMA, Yuko
; APPLICANT: KUMAGAI, Hiromichi
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
; FILE REFERENCE: 0059-1142-0PCT
; CURRENT APPLICATION NUMBER: US/09/331.581
; EARLIER FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: PCT/JP98/04929
; EARLIER FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: JP 9-314608
; EARLIER FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 24
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-331-581-24

Query Match 11.7%; Score 333.5; DB 4; Length 337;
Best Local Similarity 30.0%; Pred. No. 2.7e-22;
Matches 105; Conservative 55; Mismatches 139; Indels 51; Gaps 16;

Qy 24 GGAVPPAARASAPGSLRAVYHMTPPSGWLCDDP-----QRPVTHGAYQIYLYHSDQNG 77
Db 10 GFAAKMSASMTNETSDRPLVHFTPNKGMNDPGLWYDEKDAKWHYFQ--YNPNDTVMG 67
Qy 78 -PGGDHASTTDGVAFTHHG-TVMPLRDPFPVWGSAAVVGTAAGFAGAV-----VA 129
Db 68 TPLFWGHATSDDLTWNEDQPIAIPKRNDSGAFSGMSWVVDYNNYNTSGFFNDTIDPQRCVA 127
Qy 130 LATQPTDGRVKYQBYLYWSTDGGFTFTAL-PDPVIVNTDGRAATTPAEIENAEWFRDPK 188
Db 128 IWTYNTP---ESEQYISYSTDGGYTFTEYQKNPVLA-----ANSTQFROPK 171
Qy 189 IHWDTARGEWCVICRLR--YAAFYTSNLRDWTLRNFDYPNHALGG--IECPDLFEIT 244
Db 172 VEWYEPSQKWIMTAASKSQDYKIEIYSSDDLKSWKTESAF--ANEGLFGYQYECPLIEVP 229
Qy 245 A--DDGTRHVVLAASMDAYGIGLPMTYA---YWTGTWDGEOFHADDLPQWLDGWKNYA 299
Db 230 TEQDPSKSYWWMFISINP---GAPAGGSFNQYFVGSFNGTHEFAFDNOSRVVDFCKDYA 286
Qy 300 AVTVPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGONSIVRELRL 349
Db 287 LOTFTNTDPTVGSALGIAWASNWEISA-FVTPNP---WRSSMSLVKRFSL 332
```

RESULT 6

US-09-331-581-23
; Sequence 23, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
; APPLICANT: TOHDA, Hideki
; APPLICANT: HAMA, Yuko
; APPLICANT: KUMAGAI, Hiromichi
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
; FILE REFERENCE: 0059-1142-OPCT
; CURRENT APPLICATION NUMBER: US/09/331,581
; CURRENT FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: PCT/JP98/04929
; EARLIER FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: JP 9-314608
; EARLIER FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 23
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Schwanniomyces occidentalis
US-09-331-581-23

Query Match 10.2%; Score 291; DB 4; Length 332;

Best Local Similarity 28.1%; Pred. No. 2e-18;
Matches 94; Conservative 58; Mismatches 125; Indels 58; Gaps 17;

QY 40 RAVYHMTTPSGWLCDPQRPV--TTHGAYQLYLHSDQNG---PGGWDHASTIDGVAFTH 94
DB 26 RPLHHTPEKGMNDPNSTGYDRTAKTWHLYFOYNPNATAWGOPLYWGHATSNDLVHWE 85
QY 95 HG-TVMPLRPDPFVWGSAAVVGTAAGCAGA-----VVALATQPTDGVKRYQOYL 147
DB 86 HENAIGPEHDNEIGFSGSIVVDHNTSGFNSIDPNQRIAYI---TNNMPLQTDIA 142
QY 148 WSTDGQFTTALPDPVIVNTDGRAATPAEIAENAEWFRPKIHW-----DTARGEWC 200
DB 143 FSLDGGYTETKYENNVIDV-----SSNQFRPKVFWHERFKSMDHCSE--- 187
QY 201 VIGRLRYAIFYSPNLRDWTLRNFD--YPNHALGIECPDLFEITADGTR--HWVLAAS 257
DB 188 -TARVRIQIF-GSANLKNVLNNSFSSGYGNQYG---MSRLIEVPIENSDKSKWYMF 242
QY 258 MDAYGTGLPM---TYAYMTGTWDEGFHADDLPQWLDGWDWYAAVTPSIDAPETKRL 314
DB 243 INP---GSLPGSSINQYFVGDFDGFQFVDDSQTRFVDIGKDFYARQTFSEV---EHGVL 296
QY 315 ATAMNNKRYARDVPTDASDGYNGONSIVRELRL 349
DB 297 GLAWASNQYADQ-VPTNP---WRSSTSLARNYTL 327

RESULT 7

US-08-245-809-1
; Sequence 1, Application US/08245809
; Patent No. 5665579
; GENERAL INFORMATION:

; APPLICANT: Fitzmaurice, Leona
; APPLICANT: Mirkov, Erik
; APPLICANT: Butler, William O.
; APPLICANT: Dickinson, Craig D.
; APPLICANT: Elliott, Kathryn J.
; APPLICANT: Konno, Yoshihiro
; TITLE OF INVENTION: NOVEL INVERTASE GENE(S) AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Fitch, Even, Tabin & Flannery
; STREET: 135 South LaSalle Street, Suite 900
; CITY: Chicago
; STATE: IL

; COUNTRY: U.S.A.
; ZIP: .60603-4277
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/245,809
; FILING DATE: 17-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/771,331
; FILING DATE: 03-OCT-1991
; APPLICATION NUMBER: US 07/660,344
; FILING DATE: 22-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 51514
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-552-1311
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-245-809-1

Query Match 9.7%; Score 275; DB 1; Length 635;

Best Local Similarity 24.2%; Pred. No. 1.5e-16;
Matches 137; Conservative 76; Mismatches 236; Indels 118; Gaps 28;

QY 2 TPAISRRAYLQAGACALALIFGAVPPAARASAPGS-LRAVYHMTTPSGWLCDPQRPVT 60
DB 65 STAPPGRGVSQGVSDKTFRDVAGASHVSAWSNMLSWORTAYHFQPKNNMNDPNGLY 124
QY 61 THGAYQLYLHSDQNN--GPGGWDHASTIDGVAFTHHGTVP--LRPD-----FPWWSG 111
DB 125 HKGWYHLFYQNPDSAIWGNITWGHAVSKDLI---HWLYLPFAMVPDQWYDINGVMTGS 180
QY 112 AVVGTANTAGGAGAVVALATQPTDGVKRYQOYLWSTDGQFTTALDPDVI--VNTDG 169
DB 181 AT------LPDQGLMMLYTGTDYD--VQVQNLAIPAN-----LSDPLLDWYKFKG 224
QY 170 RAATTPAEIENAEWFRPKIHW--DTARGEWCV---IGRLRYAIFYSPNLRDWTLLRN 224
DB 225 NPVLVPPPGIGVKDFRDPITANTGPONGWLLTIGSKIKGTGVALVYETSNTFSKL--- 281
QY 225 FDYPNHALGGI---ECPDLFEITA-----DDGTRHWVLAASM-----DAYGIGL 265
DB 282 LDGVLHAVPCTGMWECVDFYVSTKKTNGLDTSYNGPGVKH-VLKASLDDNKQDHVAIG- 339
QY 266 PMTYAYWTGTWDEGFHADDLPQWLDGWDWYAAVTPSIDAPETKRLAIAAMNNKYA 325
DB 340 --TYDLGKNKWTDPNPELDCGIGLRIDYG-KYASKTF--YDPKKERRVLMGWIGETDSE 394
QY 326 ARDVPPTDASDGYNGONSIVRELRLARPGGWYTLTSPVAALTNVYVTTATTLTLDRTVD- 383
DB 395 SADL-----QKGNASVQSIPRTVLYDKKTG--THLLOPVEETESLRVGDPV--KQVDLQ 446
QY 384 -GSVLPWNGRAYEIELDIAWDT-----ATNVGISVSGSPDGTGTRHTNIGKYG--- 429
DB 447 PGSIELLRVDSAAELDIEASFVKVALQGIIEADHVGFCSTSGCAASRGILGPGFVIV 506
QY 430 -AD-----LYVDRG-----PSDLAGYSLAPYSRAAAPIDPGARSV 463
DB 507 IADQTLSELTPVYFYISKAGADRAETHFCADQTRSEAFPGVKQYGVSSVPVLDGKHS- 565
QY 464 HLRILVDTQSVEFVNAGHTVLSOOVH 490
DB 566 -MRLLDVHSIVESFAOQGRVTITSRIY 591

RESULT 8
US-08-296-624-2
; Sequence 2, Application US/08296624
; Patent No. 5658773
; GENERAL INFORMATION:
; APPLICANT: Bennett, Alan B.
; APPLICANT: Klann, Ellen
; TITLE OF INVENTION: Tomato Acid Invertase Gene
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth A. Weber
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,624
; FILING DATE: 26-AUG-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/770,970
; FILING DATE: 07-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 2307B-036510US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 636 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-296-624-2

Query Match 9.6%; Score 274.5; DB 1; Length 636;
Best Local Similarity 24.1%; Pred. No. 1.6e-16;
Matches 137; Conservative 76; Mismatches 236; Indels 119; Gaps 28;
Qy 2 TPAISRRVLOGAGAGALALIFGGAVPPAARASAFGS-LRAVYHMTPPSGWLCDDPQRPVT 60
Db 65 SPAPSRGVSQGVSDKTPRDVAGASHVSYANSNAMLQWRTAYHFQPKQKNMNDPNGLPY 124
Qy 61 THGAYQLYLHSDQNN--GPGGWDHASTTDGVAFTHHCTVMP--LRPD-----FPVWSGS 111
Db 125 HKGWTHLFYQYNPDSAINWNTTGWHAVSKDLI---HWLYLFFAMVPQWYDINGVWTGS 180
Qy 112 AVVGTTANTAGFGAGAVVALATQPTDGVKRYOEXLYWSTDGFTTALPDPI---VNTD 168
Db 181 ATI-----LPDGOIMLYTGDDTDY--VQVONLAYPAN-----LSDPLLLLDWVKEK 224
Qy 169 GRAATTPTAEIENAEWFRPKIHW-DTARGEWCV-----IGRLRYAAFTSPNLRDWTLLR 223
Db 225 GNPVLVPPPGIGVDFRDPPTTAWTGSPQNGWLLTIGSKIGTGVVALVYTSNFTSKL-- 282
Qy 224 NFDPNHALGI---ECPLDFEITA-----DDGTRHWLAASM-----DAYGIG 264
Db 283 -LDGVLHAVPGTGMWECVDFYFVSTKKTNGLTDSTYNGFGVKH-VLKASLDDNKQDHYAIG 340
Qy 265 LPMYAYWTGWDGQFQHADDLTPOWLDGWDWVAATWPSIDAPETKRLAIAMNNKY 324
Db 341 ---TYDLGKNKWTDPNPELDCGIGLRLDYG-KYYASKTF--YDPKRRVRLWGWIGTETS 394

Qy 325 AARDVPTDASDYGNGNSIVRELRLARQPGWYTWLLSTPVAALTYVTATTLTPORTVD- 383
Db 395 ESADL---QKGWASVQSIPRTVLVDKKTG---THLLQWPVEEIESLRVGDPTV--KQVDL 446
Qy 384 --GSVLPWNGRAYEIELDIAWDT-----ATNVGISVGRSPDGTGRHTNICKYG-- 429
Db 447 QPGSIELLRVDSAAELDIEASFVDKVALQGIIEADHVGFSCSTSGGAASRGILPGFVI 506
Qy 430 --AD-----LYVDRG-----PSDLAGYSLAPYSRAAAPIDPGARS 462
Db 507 VIADQTLSELTPVYFYISKGADGRAETHFCADQTRSSSEAPGVGKQVYGVSSVPVLGKHS 566
Qy 463 VHLRILVDTQSVVEFVNAGHTVLSQQVH 490
Db 567 --MRLLDVHSIVESFAQGRRTVITSRIY 592
RESULT 9
US-09-134-001C-3895
; Sequence 3895, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-3895

Query Match 8.3%; Score 235.5; DB 4; Length 492;
Best Local Similarity 23.5%; Pred. No. 4e-13;
Matches 123; Conservative 69; Mismatches 202; Indels 129; Gaps 24;
Qy 40 RAVYHMTPPSGWLCDDPQRPVTHGAYQLYLHS-----DONNGPGGWDHASTTDGVAFTHH 95
Db 35 RQTFHIQPEMGLLNDPENGLIYNGHY--YISHQWFLPGAVHGLKYWFNYKSKDLLHFEQ 92
Qy 96 GTVMPLRPD-----FPVWSGSVAVGTANTAGFGAGAVVALATQPTDGVKRYOEQ--YLYW 148
Db 93 GTL--LKPDTKYDSHGVYSGSAF-----EYQNHLYMYT 124
Qy 149 STDGGFTTALPDPI--VNTDGRAATPAETENAE-----WFRDPKTHMDTARGEWC 200
Db 125 GNRDQHNRRISSOMIARNKDKTEKPKPVHQPEGYTSFHRDPKV--EKNLSQLYA 182
Qy 201 VTG-----RLRYAFTYSPNLRDWTLLRRNFDYPNHALGI-----ECPLDFEIT 244
Db 183 ILGAQNEENEMGLLYRSQDVVDW-----HFEGEIKTNLQFGYMWCEPDYFRLS 232
Qy 245 ADDGTRHWVLAASMDAYGIGLPMY--AYWCTWDGEGFHADDLTPQWLDGWDWYAAVT 302
Db 233 NKDVT--LMCPQGVAEQDKFRNIYQSGYMGDLNFFNLFDFHESFQELDNGDFYAPQT 290
Qy 303 WPSIDAPETKRLAIAMNNWKYAADVPTDASDYGNGNSIVRELRLARQPGWYTWLLST 362
Db 291 F--VDA-DGQRLIGWGMG---LPDTEYPTD-KEGWAHCLTIPRVLTENG-----KLQR 338
Qy 363 PYAAL-----TNVTATTLTPDRTVDGSA-----VLPWNGRAYEIELDIATNATNVGIS 412
Db 339 PKQLEDLRTNKETAL-----GYANKFKRKLHPYEGKQYEMIIDILENDASEIYFE 389
Qy 413 VGRSPDGTGRHTNICKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARS-----HL 465

Db 390 LRSSRSESLITLTKHKNKLTLDRTDS-----GTLPSNVDTGTRSTRILDSPKLQ 439
QY 466 RILVDTSQSEVFNAGHTVLSQOVHFAEGDTGISLYTDGGPAH 508
Db 440 QIPVDTSIEIFCNDGERVLTSLRIFPNEDATGKASTESQVY 482

RESULT 10

US-08-860-091A-2

; Sequence 2, Application US/08860091A

; Patent No. 6057494

; GENERAL INFORMATION:

; APPLICANT: Koops, Dr. Andries Jurriaan

; APPLICANT: Van der meer, Dr. Ingrid Maria

; APPLICANT: Van Tunen, Dr. Arjen Johannes

; TITLE OF INVENTION: DNA sequences encoding carbohydrate

; TITLE OF INVENTION: polymer synthesizing enzymes and method for producing transgen

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LADAS & PARRY

; STREET: 26 WEST 61 STREET

; CITY: NEW YORK

; STATE: NY

; ZIP: 10023

; COUNTRY: USA

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3 1/4" Floppy disk 1.44MB

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: Microsoft Windows for Workgroups

; SOFTWARE: WordPerfect 8.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/860,091A

; FILING DATE: 02-JULY-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/NL/00012

; FILING DATE: 08-JANUARY-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Cord, Janet I.

; REGISTRATION NUMBER: 33,778

; REFERENCE/DOCKET NUMBER: U-011329-3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 708-1935

; TELEFAX: (212) 246-8959

; TELEX: 233288

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 630 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-860-091A-2

Query Match 7.2%; Score 205.5; DB 3; Length 630;
Best Local Similarity 21.2%; Pred. No. 3.1e-10;
Matches 124; Conservative 77; Mismatches 202; Indels 183; Gaps 29;

QY 20 ALIFGAVPPAARASAPGSL-----RAYVHMTPPSGWLCDDPQRPVTHGAYQLYYL 70

Db 65 ATTFVTQLPNIDLKRVPGKLDSSAEVEWQSTYHFQPKNFISDPDGPVHMGWYHLFYQ 124

QY 71 HSDONN--GPGGDHASTTGDVAFTHHGTVMPP--LRPDPFVWSSGSAVGTANTAGFCAGA 126

Db 125 YNPQSAIWNITWGHVSVDKMINWFH----LPFAMVPDH--WYDIEGVTGSAVLPNGQ 178

QY 127 VVALATQPTDGVKRYQEOYLWSTGDTGFTTALPDPIV---NTDGRAATTAEIENAEW 183

Db 179 IIMLYS--GNAYDLSOVQCLAYAVNSS-----DPLLEWKYEGNPVLLPPPGVGYKD 229

QY 184 FRDPKIHMDTARGEWCVCIG-----RLRYAAFTYSPNLRDWTLRNFDYPNHALGEGIECP 238

Db 230 FRDPSTLWSPGDGEYRWVNGSKHNETIGCALLYHTTNTFHFELKEEVLHAPHTGMWECV 289

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 184 FRDPKIHMDTARGEWCVCIG-----RLRYAAFTYSPNLRDWTLRNFDYPNHALGEGIECP 238

Db 230 FRDPSTLWSPGDGEYRWVNGSKHNETIGCALLYHTTNTFHFELKEEVLHAPHTGMWECV 289

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263 IGLPMTYAYWTGTDGEOQFHAD-----DLTPQWLDGWDWYAAVTWP 304

Db 350 IGL-----RYDFGKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262

Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLSQSGDEDRHDWYAIGSYDIVNDKWYPDDPENDVG 349

QY 263

Qy 305 S---IDAPETKRLAIW-----MNNWKY-AARDVPTDASDGYNGONSIVRELRAROP 353
Db 401 RTVVLDL-ETKTLNLIOWPIEETENLRSSKKYDEKDV-----ELR----P 439
Qy 354 GGWYLLSTPVAALTNV-VTATTLTLPDRTVDSAVLPWNGRAYEIELDIANDTATNNGIS 412
Db 440 G---ALVPLETGTATQLDIVATFEIDQKMLEST-----LEADVLENCTTSEG-S 484
Qy 413 VGRSPDGRTRHNTNIGRYGADLVXDRGPS-----DLAG-----YSLAPYSR----- 451
Db 485 VARSV-----LGPFGVVLADAQRSEQLPVYFVIANDIDGTSRTYFCADETRSSKDV 537
Qy 452 -----AAAPIDGARSVHLRLVDTSVEFVNAGHTVLSQQVH 490
Db 538 VGWVYGSVPVLPGEK-YNMRLLVHDHSIVEGFAQNGRTVVTSRVY 582

RESULT 12

US-08-860-091A-4
; Sequence 4, Application US/08860091A
; Patent No. 6057494

GENERAL INFORMATION:

APPLICANT: Koops, Dr. Andries Jurriaan
APPLICANT: Van der meer, Dr. Ingrid Maria
APPLICANT: Van Tunen, Dr. Arjen Johannes

TITLE OF INVENTION: DNA sequences encoding carbohydrate

TITLE OF INVENTION: polymer synthesizing enzymes and method for producing transgene

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: LADAS & PARRY

STREET: 26 WEST 61 STREET

CITY: NEW YORK

STATE: NY

ZIP: 10023

COUNTRY: USA

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/4" Floppy disk 1.44MB

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Microsoft Windows for Workgroups

SOFTWARE: WordPerfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/860,091A

FILING DATE: 02-JULY-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/NL/00012

FILING DATE: 08-JANUARY-1996

ATTORNEY/AGENT INFORMATION:

NAME: Cord, Janet I.

REGISTRATION NUMBER: 33,778

REFERENCE/DOCKET NUMBER: 0-011329-3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 708-1935

TELEFAX: (212) 246-8959

TELEX: 233288

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 615 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-860-091A-4

Query Match 6.6%; Score 189; DB 3; Length 615;

Best Local Similarity 20.0%; Pred. No. 9.4e-09;

Matches 115; Conservative 70; Mismatches 179; Indels 210; Gaps 27;

Qy 40 RAVYHMTPPSGWLCDPQRPVTHGAYQLYLHSDQNN--GPGGWDHASTTDCGVAFTHGT 97
Db 88 RTAFHFQPAKFIYDPDQGLFHMGMWYHMFYQYNPYAPVWGNMWSGHSVKDMINWYE--- 144
Qy 98 VMLPRDPFVWSSGSAVGTANTAGGAGAVVALATQPTDGVKRYQEQVLYWSTDGGFTFT 157
Db 145 -LPVAVPTWEVDIEGVLGGSTTVLPNGQIFALYTGNAFDSOLO-----CK 190

Qy 158 ALP-----DPVIVN-----TDGRAATTPAETENAEWERDPKIHWDTARGEVWCVIGRLR--- 206
Db 191 AVPNVNSDPLLEWVKYEDNPILYTPPGI-GLKDYRDPSTVWTGPDGKHRMIMGTKRGNT 249
Qy 207 --YAAFYTSPNLRDWTLRNFDYPNHALGGI---BCPDLFETADDGTRHWWVLAASMDAY 261
Db 250 GMLVYYVYT---DYNTYELLDEPLHSPNTDMWECVDFYPVSLTND---ALDMAAY 300
Qy 262 GIGLPMYAYWGTGWDGEQFHADDLTQPWLDMGWDMYAAVTWPSI-----DAPE----- 310
Db 301 GSGIKHVI---KESWEGH-----GMDWYSIGTYDAINDKWTDPNPELDVGI 343
Qy 311 -----TKRLAIAMNNWYKAAARDVPTDASDGYNGONSIVRELRARQ 352
Db 344 GLRCDYGRFFASKSLYDPLKKRRITW-----GYVGESDSA-DQDLSR- 384
Qy 353 PGGWYLLSTPVAALTNVVTATTLTLPDRTVDSGSAVLPH-----NGRAYE----- 396
Db 385 --GWATV-----YNVGRTIVLDRKT-GTHLLHMPVEEVESLRYNGOEFKEIKLEP 431
Qy 397 ---IELDIADWTATNVGI-----SVGRSPDGRHHTNIGKY 428
Db 432 GSIILPDIG--TATQIDIVATFEVDQAAALNATSETDDIYGCTTSLGAAQRG---SLGPF 485
Qy 429 GADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSYH-----464
Db 486 GLAVLADGTLSELT---PVYFYIAKKADGGV-STHFCTDKLRSSILDYDGERVYGGTVP 540
Qy 465 -----LRLIVDTQSVEFVNAGHTVLSQQVH 490
Db 541 VLDDEELTMLRLVHDHSIVEGFAQNGRTVITSRAY 574

RESULT 13

US-09-245-323A-8

; Sequence 8, Application US/09245323A

; Patent No. 6365800

GENERAL INFORMATION:

APPLICANT: Caimi, Perry G.

TITLE OF INVENTION: Transgenic Crops Accumulating Fructose Polymers and

TITLE OF INVENTION: Methods for Their Production

FILE REFERENCE: BBI082 US NA

CURRENT APPLICATION NUMBER: US/09/245,323A

CURRENT FILING DATE: 1999-02-05

PRIOR APPLICATION NUMBER: 60/077,727

PRIOR FILING DATE: 1998-03-12

NUMBER OF SEQ ID NOS: 8

SOFTWARE: MICROSOFT OFFICE 97

SEQ ID NO 8

LENGTH: 615

TYPE: PRT

ORGANISM: Helianthus tuberosus

US-09-245-323A-8

Query Match 6.6%; Score 189; DB 4; Length 615;

Best Local Similarity 20.0%; Pred. No. 9.4e-09;

Matches 115; Conservative 70; Mismatches 179; Indels 210; Gaps 27;

Qy 40 RAVYHMTPPSGWLCDPQRPVTHGAYQLYLHSDQNN--GPGGWDHASTTDCGVAFTHGT 97
Db 88 RTAFHFQPAKFIYDPDQGLFHMGMWYHMFYQYNPYAPVWGNMWSGHSVKDMINWYE--- 144
Qy 98 VMLPRDPFVWSSGSAVGTANTAGGAGAVVALATQPTDGVKRYQEQVLYWSTDGGFTFT 157
Db 145 -LPVAVPTWEVDIEGVLGGSTTVLPNGQIFALYTGNAFDSOLO-----CK 190
Qy 158 ALP-----DPVIVN-----TDGRAATTPAETENAEWERDPKIHWDTARGEVWCVIGRLR--- 206
Db 191 AVPNVNSDPLLEWVKYEDNPILYTPPGI-GLKDYRDPSTVWTGPDGKHRMIMGTKRGNT 249
Qy 207 --YAAFYTSPNLRDWTLRNFDYPNHALGGI---BCPDLFETADDGTRHWWVLAASMDAY 261

Db 250 GMVLVYTT-----DYTNVELLDPLHSPVNTDMWECVDFYVNSLTNDS-----ALDMAAY 300
QY 262 GIGLPMYAYWTGDEGFHADDLTPOWLDGWDWYAAVTPSI-----DAPE----- 310
Db 301 GSGIKHVI---KESWEGH-----GMDWYSIGTYDAINDKWTDPNDELVDGI 343
QY 311 -----TKRLAIAMNNWKYAAARDVPTDASGYNGQNSIVRELRLAQ 352
Db 344 GLRCDYGRFFASKSLYDPLKKRITW-----GYVGESDSA-DQDLRS- 384
QY 353 PGGWYLLSTPVAALNYVTATTLPLDRVDSAVLPW-----NCRAYE----- 396
Db 385 -GWATV-----YNGRTIVLDKRT-GTHLLHWPVEEVESRYNGQEKETKLEP 431
QY 397 ---IELDIAWDATNNGI-----SVGRSPDGTNRHTNIGKY 428
Db 432 GSIIPLDIG-TATQLDIVATEVDQAALNATSETDDIYCTISLGAORG-----SLGPF 485
QY 429 GADLYVDRGPSLAGYSLAPYSRAAIPDPGARSVH----- 464
Db 486 GLAVLADGTLSELST-----PVYFYIAKKADGV-STHFCTDKLRSLDYGGERVYGGTVP 540
QY 465 -----LRLVLDTSQVEVFNAGHTVLSQVH 490
Db 541 VLDDEELTRMLLDHSHVSEVGFAGGRTVITSRAY 574

RESULT 14
US-09-019-385-2
; Sequence 2, Application US/09019385
; Patent No. 6147280
; GENERAL INFORMATION:
; APPLICANT: Smeekens, Josephus C.M.
; APPLICANT: Ebskamp, Michael J.M.
; APPLICANT: Geerts, Hendrikus A.M.
; APPLICANT: Weisbeek, Petrus J.
; TITLE OF INVENTION: Production of Oligosaccharides in
; TITLE OF INVENTION: Transgenic Plants
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson & KindnessPLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,385
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,470
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NL 1000064
; FILING DATE: 05-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Brogerick, Thomas F.
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: ARNO18589
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-224-0709
; TELEFAX: 206-224-0779
; TELEX: 4938023
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 626 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-019-385-2

Query Match 6.0%; Score 169.5; DB 4; Length 626;
Best Local Similarity 21.8%; Pred. No. 5.7e-07;
Matches 136; Conservative 76; Mismatches 235; Indels 177; Gaps 36;

QY 16 AGALALIFGAV-----PPAARASAPGS-----LRAVYHMTPPSGWLCDPQR 57
Db 40 ASAMAVVVVGATLLAGLRMEQAVDEEAAGGFPWSNEMQLQWQSGYHFOFTAKNYSMDPNG 99
QY 58 PVTTHGAYOLYLLHSDQNNPG-GWDHASTDTCVAETH-----HGTWMLRPDPFVW- 108
Db 100 LMYRYGWYHMFY---QYNPVGTDWD-----DCMENGHAVSRNLVQWRTLPILAMVADQWY 150
QY 109 -----SGSAVV---CTA---NTAGFGAGAV---VALATOPTDGV---RKQEOYLYW 148
Db 151 DILGVLGSGMTVLPNGTVIMITYGATNASAVEQCITATPADPNPDLRLRWTKHPANPVIW 210
QY 149 STDGGFTFTALPDPIVINTDGRAATTPAETENAEWFRDPKIH-W-DTARGEWVCVIGRL-- 205
Db 211 SPPG-----VGTKD-----FRDPTAWYDESDETWRTLLGSKDD 244
QY 206 -----RYAFTYSPNLRDWTLRNFDYPNHALGGTECPDLFEI---TADDGTR--HWVL 254
Db 245 HDGHHDGIAMMYKTKDFLNVELIPGILHRVVRTGEWECIDFYVGRRSSDNLSEMLH-VL 303
QY 255 AASM-----DAYGIGLPMYAYWTGWDGEQPHADDLTPOWLDWG---WDW---YAAVT 302
Db 304 KASMDDERHDYSLG---TYDSAANTW-----TPIDPE-LDLGIGLRYDMGKFTASTS 352
QY 303 WPSIDAPETKRLAIAMNNWKYAADVPTDASGYNGQNSIVRELRLARQPGGWYTLST 362
Db 353 F--YDPAKNRRVLMGVYGEVDSKRDV---VKGWASIQSVPTVALDEKTR--TNLLW 404
QY 363 PVAALTNVYVTTPLDRTVD-GSAV-LPWNGRAY-----EIELDIAWDATN---VGI 411
Db 405 PVETETLRNLNATELTDVTINTGTSVTHIPLRQGTTHARHAEASFHLDASAVALNEADVY 464
QY 412 SVGRSPDGTNRHTNIGYGADLYVDRGPSLAGYSLAPYSRAAIPDPG----- 459
Db 465 NCSSSGGAVNRGALGPFGL-LVLAAG--DRGEQTAIVFYVSKGLDGLGHTSFCODELRS 521
QY 460 -----ARSHRLILYDTSQVEVFNAGHTVLSQVHFAEG---DTGI 498
Db 522 SRAKDVTKRIVGSTVPVLDGEALSMRVLVDHSIVQGFDMGRTTMTSRVYPMESYOEARV 581
QY 499 SLYTDGGPAHFTG--IVVREIGQA 520
Db 582 YLFNNTGASVTAERLVVHEMDSA 605

RESULT 15
US-09-142-623-11
; Sequence 11, Application US/09142623
; Patent No. 6337201
; GENERAL INFORMATION:
; APPLICANT: Koji YANAI et al.
; TITLE OF INVENTION: -FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF
; TITLE OF INVENTION: ISOLATING -FRUCTOFURANOSIDASE GENE, SYSTEM FOR PRODUCING
; TITLE OF INVENTION: -FRUCTOFURANOSIDASE, AND -FRUCTOFURANOSIDASE VARIANT
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/142.623
FILING DATE: September 10, 1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: 98-0989*/LC(WMC)/144
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acid residues
TYPE: Amino acid
STRANDEDNESS: No. 6337201 relevant
TOPOLOGY: No. 6337201 relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Microorganism: Penicillium roqueforti IAM7254
FEATURE:
NAME/KEY: mat peptide
LOCATION: 1..565
IDENTIFICATION METHOD: E

Search completed: May 6, 2003, 12:36:32
Job time : 23.5375 secs

Query Match 5.2%; Score 147.5; DB 4; Length 565;
Best Local Similarity 21.8%; Pred. No. 4.9e-05;
Matches 126; Conservative 68; Mismatches 226; Indels 159; Gaps 29;

Qy 26 AVPPAARASAPGSL---RAYVHTPPSGWLCDP---QRPVTHGAYQLYVYVHSDQNG 77
Db 11 SAPNMLSTLANASLFKTRPRAHLPLPPSGNIGDPCGHYTDPKT--GLFHVGWLYS---G 64
Qy 78 PGWDHASTTDCGAFTHRTGTHVPLRPDPFVWMSGSAVGTAN-TAGFGAGAVV--ALATQP 134
Db 65 1SG---ATTDDL---TYKDLNPD---GAPSIAGGKNDPLSVFDGSIPIPSGIDGMP 112
Qy 135 T-----DGVRYQEQLYVSTGDTFTALPDPIVINTDGRAATTPAET 179
Db 113 TLLYTSVSLPIHWSIPYTRGSETQSLAVSYDGGHNETKL-----NQGPVPTPPFAL 165
Qy 180 NAEWFRDPKIH---DTARGEWCYVIG-----RLRYAAYTSPNLRD 218
Db 166 NTAFRDPYVFQSPILDKSVNSTGTWYVAISGGVHVGPCQFLYRONDADFQYWEYLGO 225
Qy 219 W---TLRNFDPNHALG---GIECPDLFEITADDGTRHWVLAASMDAYGIGLP----- 266
Db 226 WKKEPLNTTWKGDWAGWGWFENFVGNVFSNAGYSEDEGEIFITLGAEGSGLPVPOVS 285
Qy 267 ---MTYAYWTGTDWGEFHADDLPQWLDWGDWYAAV--TWPSIDAPETKR-----L 314
Db 286 SIRDMLWVTGNVNDGVSFTEKPTMAGV-LDWGSAYAAAGKILPASSQASTKSGAPDRFI 344
Qy 315 AIAWNNHNY-AARDVPTDSDGNGQNSIVREL-----RLARQ 352
Db 345 SYVMTGLDFEQVGFPF-AQONMTGALLPRELNVRTISNVVDNELSRESLTSWRVARE 403
Qy 353 PGWYTLTSTPVAALNTVNTATT-----LPDRTVDGSAVLPN---GRAVEIELDIW 403
Db 404 DSGQIDLETMGISISRETYISALTSGSSPVEGKTLNAGAVFPNTSPSKFVLTAISF 463
Qy 404 DT-ATNVGISVGRSPDGRHTNICKY-----GADLYVDRGSDLAGYSLAPYSRAAPI-- 456
Db 464 PTSARDSGIQAGFVLSLSLESTTIYQFSNESIIIVDRSNTSAA-----ARTTAGILS 516
Qy 457 --DPG-----ARSVHLRLVDTQSVFVFN 479
Db 517 DNEAGRLRFLVLRNGKEQVETLELTIVVDNSVLEVYAN 555

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:34:52 ; Search time 17.5337 Seconds
(without alignments)
2563.947 Million cell updates/sec

Title: US-09-868-328B-1

Perfect score: 2845
Sequence: 1 MTPAISRAVLQAGAGALA.....TDGGPAHPTGIVREIGQAI 521

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328255 seqs, 86286685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	499.5	17.6	943	10	US-09-969-362-5
2	468.5	16.5	1277	10	US-09-969-362-3
3	461	16.2	923	10	US-09-969-362-1
4	221	7.8	635	9	US-10-003-392-14
5	206	7.2	630	9	US-10-003-392-16
6	205.5	7.2	630	9	US-10-003-392-18
7	202.5	7.1	433	9	US-09-738-626-6405
8	189	6.6	615	9	US-10-003-392-17
9	184	6.5	608	9	US-10-003-392-2
10	176	6.2	625	9	US-10-003-392-21
11	160	5.6	600	9	US-10-003-392-20
12	155	5.4	609	9	US-10-003-392-4
13	147.5	5.2	565	9	US-09-990-385-11
14	141.5	5.0	635	9	US-09-990-385-1
15	140.5	4.9	574	9	US-09-990-385-13
16	133.5	4.7	495	9	US-10-003-392-8
17	131.5	4.6	498	9	US-09-738-626-4737
18	126	4.4	390	9	US-10-003-392-6
19	126	4.4	471	9	US-10-003-392-10

20	117	4.1	451	9	US-09-712-363-171	Sequence 171, Appl
21	115	4.0	3739	9	US-09-860-846-33	Sequence 33, Appl
22	115	4.0	3739	9	US-09-988-384B-33	Sequence 33, Appl
23	115	4.0	3739	10	US-09-861-289-33	Sequence 33, Appl
24	115	4.0	11877	9	US-09-860-846-6	Sequence 6, Appl
25	115	4.0	11877	10	US-09-861-289-6	Sequence 6, Appl
26	115	4.0	12199	9	US-09-988-384B-6	Sequence 6, Appl
27	113	4.0	296	10	US-09-960-472-1	Sequence 1, Appl
28	112	3.9	666	10	US-09-771-161A-136	Sequence 136, Appl
29	111	3.9	3472	9	US-10-027-806-4	Sequence 4, Appl
30	111	3.9	3472	9	US-10-034-623-4	Sequence 4, Appl
31	111	3.9	3472	9	US-10-027-801-4	Sequence 4, Appl
32	110.5	3.9	983	10	US-09-771-161A-227	Sequence 227, Appl
33	107	3.8	682	10	US-09-798-791-2	Sequence 2, Appl
34	106.5	3.7	1352	9	US-09-784-554B-2	Sequence 2, Appl
35	106	3.7	3519	9	US-09-808-880-4	Sequence 4, Appl
36	106	3.7	4613	9	US-09-860-846-31	Sequence 31, Appl
37	106	3.7	4613	9	US-09-988-384B-31	Sequence 31, Appl
38	106	3.7	4613	10	US-09-861-289-31	Sequence 31, Appl
39	105.5	3.7	811	9	US-10-011-366-7	Sequence 7, Appl
40	105.5	3.7	2710	9	US-10-011-366-6	Sequence 6, Appl
41	105	3.7	742	9	US-10-077-111-11	Sequence 11, Appl
42	105	3.7	1344	9	US-09-738-626-6888	Sequence 6888, Ap
43	104	3.7	655	9	US-10-124-880-14	Sequence 14, Appl
44	103.5	3.6	559	10	US-09-725-735A-15	Sequence 15, Appl
45	103.5	3.6	630	10	US-09-725-735A-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-969-362-5
; Sequence 5, Application US/09969362
; Patent No. US20020076790A1
; GENERAL INFORMATION:
; APPLICANT: Moller, Soren
; APPLICANT: Johansen, Charlotte
; APPLICANT: Schafer, Thomas
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Hoeck, Lisbeth Hedegaard
; TITLE OF INVENTION: A 2,6, b-D-Fructan Hydrolase Enzyme And
; FILE REFERENCE: 5540.200-US
; CURRENT APPLICATION NUMBER: US/09/969,362
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 09/397,885
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: PA 1998 01623
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: 60/101,615
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/111,675
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Paenibacillus macerans
US-09-969-362-5

Query Match	17.6%	Score 499.5;	DB 10;	Length 943;
Best Local Similarity	30.8%	Pred. No. 7e-33;		
Matches 157;	Conservative 67;	Mismatches 223;	Indels 63;	Gaps 19;
QY	40	RAVYHMTPPSGWLCDPQRPVTHGAYQLYLSHQDQNGPGWDHASTTDCGVAFTHTGTMV 99		
Db	383	RPDYHYPARGASDSDNGLVYVEGYHLF--HQD----GGTWAHAVSTDLVHKRLPIAL 436		
QY	100	PLRPDPVWVGSAVGTANTAGF-----GAGAVVALATQTDGVRKYQEQYLWSTDGFG 154		
Db	437	PWDLGHVWVGSAVADLHNASGLFADSGKGLIAYTTSNPDRPNGNQRIGLAYSKDGR 496		

Db 446 PWNHGHVWSGSAVADMTNAGSLFGDSGKGLIAYTTFNPDSPNGNORIGLAYSKDOOR 505
Qy 155 TFTAALPD-PVIVNTDGRAATPAETENAEW-PROPKIHWDTARGEMVGVIGRLRAAEVY 212
Db 506 TWEYSKERPIVENPKSGN-----EAGNDFRDPKVIIRDENNRMVWVSGGDHIREYT 560
Qy 213 SPNLRDWTLRNFDYPNHALGGI-ECPDLEFETADDDGT--RHWVLAASMDAYGIGLPMW- 268
Db 561 STNLLDWTLDNWGYDGVYRGVWECPLFQLPV-DGTSQKKWMMISTGAN---PKTG 615
Qy 269 ---YAYWTG--TWGEQPHADDLTPQWL--DWGWDWYAAVTPSIDAPETKRLAIAMNWN 321
Db 616 GSDAEYFYGHLTADG-KFVNNDNPAGKVLRTDFGKFEYASMSFANN--PDHRTVMAMWNTN 672
Qy 322 WKYAADRVPTDASDYGNGSIVRELRLARQPGWYTLISLTPVAALTNVYATTTLPDRT 381
Db 673 WDYFPA-FPT---SNWKGELIPREVSLLVTTEDG-IRMVQPIKEL-----ESLRKPLYS 722
Qy 382 VDGSAVLPWNGR-----AYEIELDI---AWDTATNVGISVGRSPDGRHNTNIGKYGA 430
Db 723 ASKNSVSPSSGNLLKGIISGAYEIEAEIETSTVTTEFGENIREG--ANQKTVVGYKAS 780
Qy 431 D--LYVDRGPDLAGYSLAPYSRAAAPIDPGARSVHLRLVDTQSVFVFNAGHTVLSQQ 488
Db 781 DSRMFVDRTASGETDFSNLFSKKHEAPTOMENNRIMKRLVDESSVEAFGNDGKYVFSDV 840
Qy 489 VFAEGDGTGISLYTDGGPAHFTGIIVREI 517
Db 841 IFPDPAASAMSFYKGGNNVVSUKVHQL 869

RESULT 4

US-10-003-392-14
; Sequence 14, Application US/10003392
; Patent No. US20020170086A1

; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.

; APPLICANT: Calmi, Perry G.
; APPLICANT: Stoop, Johan M.

; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BBI463 US NA

; CURRENT APPLICATION NUMBER: US/10/003.392
; CURRENT FILING DATE: 2001-10-30

; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10

; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97

; SEQ ID NO 14
; LENGTH: 635

; TYPE: PRT
; ORGANISM: Parthenium argentatum Grey

US-10-003-392-14

Query Match 7.8%; Score 221; DB 9; Length 635;
Best Local Similarity 21.5%; Pred. No. 4.5e-10;
Matches 126; Conservative. 82; Mismatches 207; Indels 170; Gaps 29;

Qy 12 QGAGAGALALIFGGAVPAAASAPGSL-----RAYVHMTPPSGWGLCDPQRPVTH 62
Db 69 QGA-----ATFTTQLPKIDMKRVPGLDSDGADVQWQORSAYHFQDKNYSIDPDGPMYHM 123
Qy 63 GAYQLYLHSDQNN--GPGGDHASTTGDVAFTHHGTVM--LRPDPFVWSGSAVVGATN 118
Db 124 GWYHLFYQNPESAIWGNITWGHVSVDKMINWFH---LPFAMVPDH--WYDIEGVMTGS 177
Qy 119 TAGFGAGAVVALATOPTDGVKRYOEQLYWSSTDGGETFTALPDPIV---NTDGRAATTP 175
Db 178 ATVLNPGEIIMLYT--GNAYDLSQVQCLAYAVNSS-----DPLLEWKKEVGNPVL 228
Qy 176 AEIENAEWFRDPKIHWDTARGEWVCVIG-----RLRYAAFTYSPNLRDWTLRNFDY 230

Db 229 PPGVGYKDFRDPSTLWLGPDGEYRMVMSGKHNETIGCALIYHTTFTFELNEEVLHAVP 288
Qy 231 ALGGIECPDLPEIT-----ADGGR-HWVLAASMDA-----YGICLPWYAYWTCT 275
Db 289 HTGMMWECVDLYPVSHTHTNGLDMDVNGPNVYVVLKOSGDEDRHDWYAG-----SYD 340
Qy 276 WDGEQFHADDLTPQ-----WLDWGWDWYAAVTPSIDAPETKRLAIAMNWNKYAARD 328
Db 341 WYNDKWPYD--PENDVGIGLRYDYG-KFYASKTF-----YDQHKRRVLM----- 383
Qy 329 VPTDASDYGNGSIVRELRLARQPGWYTLISLTPVAALTNVYATTTLP----- 378
Db 384 -----GYVGEDT-PEKYDLTK--GWANILNRTVTVLDTKTKTNLIQWPIEETELR 432
Qy 379 ---DRTVD---CSAVLPWNGRAVE-----TELDIAWDTATNVGIS 412
Db 433 SKKYDKFVDELVRPGSLPLEIGTATQLDIVATFEVDQMMLLESTLEADVLFNCTTSVG-S 491
Qy 413 VGRSPDGRHNTNIGKYGADLYVDRG-----PSDLAG-----YSLAPYSRAAAPID 457
Db 492 VGRGV-----LGFPGVVVLADAQRTQLPVYFYIANDTDCGTSRTYFCADETRSSKD 544
Qy 458 PCA-----RSVHLRLVDTQSVFVFNAGHTVLSQQVH 490
Db 545 VCKWYGVSSVPVLPNEKYNMRLLDVHSIVEGFAQNGRTVVTSRVY 589

RESULT 5

US-10-003-392-16
; Sequence 16, Application US/10003392
; Patent No. US20020170086A1

; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.

; APPLICANT: Calmi, Perry G.
; APPLICANT: Stoop, Johan M.

; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BBI463 US NA

; CURRENT APPLICATION NUMBER: US/10/003.392
; CURRENT FILING DATE: 2001-10-30

; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10

; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97

; SEQ ID NO 16
; LENGTH: 630

; TYPE: PRT
; ORGANISM: Helianthus sp.

US-10-003-392-16

Query Match 7.2%; Score 206; DB 9; Length 630;
Best Local Similarity 21.8%; Pred. No. 7.7e-09;
Matches 125; Conservative 81; Mismatches 204; Indels 164; Gaps 29;

Qy 28 PPAARASA-----PGSL-----RAYVHMTPPSGWGLCDPQRPVTHHAYQL 67
Db 62 PPAATTFATQLPNIDLRKVPKGLDSSAEVWQORSAYHFQDKNFISDDPGPMYHGWYHL 121
Qy 68 YLHSDQNN--GPGGDHASTTGDVAFTHHGTVM--LRPDPFVWSGSAVVGATANTAGFG 123
Db 122 FYQYNPESAIWGNITWGHVSVDKMINWFH---LPFAMVPDH--WYDIEGVMTGSATVLP 175
Qy 124 AGAVVALATOPTDGVKRYOEQLYWSSTDGGETFTALPDPIV---NTDGRAATTPAEIEN 180
Db 176 NGQITIMLYT--GNAYDLSQVQCLAYAVNSS-----DPLLEWKKEVGNPVLFP 226
Qy 181 AEWFRDPKIHWDTARGEWVCVIG-----RLRYAAFTYSPNLRDWTLRNFDY 235
Db 227 YKDFRDPSTLWLGPDGEYRMVMSGKHNETIGCALIYHTTFTFELKEEVLHAVPHTGMW 286
Qy 236 ECPDLFEIT-----ADGGR-HWVLAASMDA-----YGICLPWYAYWTGWDGEQ 280

```
Db 287 ECVDLPVSTVHTNGLDWDVNGPNVYVLKQSGDEDRHWDYALG----SYDVVNDKWYPPD 343
QY 281 FHADDLTPOWLDGWYAAVTPSIDAPETKRLAIAMNNWKYAAARDVPTDASGYNGQ 340
Db 344 PENDVGIGLRYDFG-KFYASKTF---YDQHKRRRLW-----GYVGE 381
QY 341 NSVRELRLARQGGWYTLSTPAAL-----TNY-----VTATTLDPDRTVD----- 383
Db 382 TD-PORYDISK---GWANILNIPRTVLTDTKTNTLNQIPIETENLSRKYDEFKDV 437
QY 384 --GSAPLPNNGRAYE-----LELDIANDTATNNGISVGRSPDGTHTN 424
Db 438 RPSGLPLEIGTATQLDIVATFEIDQKMLESTLEADVLFNCTTSEG-SVARGA----- 489
QY 425 IGYGADLYVDRGSP-----DLAG-----YSLAPYSR-----AAAPI 456
Db 490 LGPGVGVVLADAQRSEQLPVYFIANDIDGTSRTYFCADETRSDKDSVSGKWYGVSSVPV 549
QY 457 DPGARSVHLRLVDTSQSVFVNAGHTVLSQQVH 490
Db 550 LPGEK-YNMRLLDHSHSVEGFAQNGRTVVTSRVY 582

RESULT 6
US-10-003-392-18
; Sequence 18, Application US/10003392
; Patent No. US20020170086A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Calmi, Perry G.
; APPLICANT: Stoop, Johan M.
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BB1463 US NA
; CURRENT APPLICATION NUMBER: US/10/003,392
; PRIOR FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 630
; TYPE: PRT
; ORGANISM: Helianthus tuberosus
US-10-003-392-18

Query Match 7.2%; Score 205.5; DB 9; Length 630;
Best Local Similarity 21.2%; Pred. No. 8.4e-09;
Matches 124; Conservative 77; Mismatches 202; Indels 183; Gaps 29;

QY 20 ALIFGCAVPPAARASAPGSL-----RAVYHMTPPSGWCLCDPQRPVTHGAYOLYYL 70
Db 65 ATTFVQLNIDLKRVPGLDSSAEVWORSYHFQDPKFNFTSDPDGPMYHGWYHLFTQ 124
QY 71 HSDQNN--GPGGDHASTTGDGVAFTHTGTVMP--LRPDPFVWMSGSAVVGTTANTAGFGAGA 126
Db 125 YNPQSAIMGNITWGHVSXDMINWFH---LPPAMVPDH--WYDIEGVTGSAVLPNQ 178
QY 127 VVALATOPTDGVKRYEQVLYSTDGGFTTALPDPIV---NTDGRATTTAEIENAEW 183
Db 179 IMLYS--GNAYDLSQVCLAVNNS-----DPLLEWKYEGNPVLLPPPGVGYKD 229
QY 184 FRDPKTHWDARGWCVTVG-----RLRYAAFTSPNLRDWTLLRRNFDPNHALGIECP 238
Db 230 FRDPSTLWSPGDEYRWMSGKHNETIGCALLYHTNTFTFELKEVLAHVPHGWEVY 289
QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YCG 262
Db 290 DLYPSTVHTNGLDVNDGPNVYVLKQSGDEDRHWDYALGSDYDINDKWYPPDENDVG 349
QY 263 IGLPWTYAWTCTWDEGEQPHAD-----DLTPQWLDWGNWDYAAVTPW 304
```

```
Db 350 IGL-----RYDGKFTYASKTFYDQHKRRRLVNGYVGETDQPKYDLKSGWANILNIP 400
QY 305 S-----IDAPETKRLAIAM-----MNNWKY-AARDVPTDASDGYNGONSVIRELRARQP 353
Db 401 RTVVLDL-ETKTNLIQWPIEETENLSRKYDEFKDV-----ELR---P 439
QY 354 GGYWYTLSTPVAALTY-VTATTLDPDRTVDGSAVLPWNGRAYEIELDIANDTATNNGIS 412
Db 440 G-----ALVPLEIGTATQLDIVATFEIDQKMLEST-----LEADVLFNCTTSEG-S 484
QY 413 VGRSPDGTHTNIGKYGADLYVDRGSP-----DLAG-----YSLAPYSR----- 451
Db 485 VARSV-----LGPGVGVVLADAQRSEQLPVYFIANDIDGTSRTYFCADETRSDKDSV 537
QY 452 -----AAAPIDPGARSVHLRLVDTSQSVFVNAGHTVLSQQVH 490
Db 538 VGKWWYGVSSVPVLPGEK-YNMRLLDHSHSVEGFAQNGRTVVTSRVY 582

RESULT 7
US-09-738-626-6405
; Sequence 6405, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6405
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6405

Query Match 7.1%; Score 202.5; DB 9; Length 433;
Best Local Similarity 24.8%; Pred. No. 9.1e-09;
Matches 132; Conservative 49; Mismatches 166; Indels 185; Gaps 32;

QY 38 SLRAVYHMTPPSGWCLCDPQRPVTHGAYOLYLSQNNNGPG-----GWDHAST--- 86
Db 11 SLRPAYHVTPPQGRNDPNMGVYDGTDLHYVYQHD-----PGFPFAPKRTGWAHTTPTL 65
QY 87 -TDGVAFTHTGTVMP--LRPDPF-----VWMSGSAVVGTTANTAGFGAGAVVALATQPTDGV 138
Db 66 GPQRLQWTH-----LPDALYPDASYDLDDGCGYSGAVFTDGTCLKLYFTGNL-----KIDGK 115
QY 139 RXYEQVLYSTD-----GSETFTALPDPIVNTDGRAA-TTPAEIENAEWFRDPKRIHWD 192
Db 116 RR-ATONLVEVEDPTCLMGIIHRRSPKNPLI---DGPASGFTP-----HYRDPMLSPD 164
QY 193 TARGE-WVCVIGRLR-----YAAFTSPNLRDWTLLRRNFY-----PNHALGGI- 235
Db 165 ---GDGKWKVLGAQRENLTGAALYLRSTDLNWEFSGEITFDLSDAQPGSAPDLVPGGYM 221
```



```

Qy 413 VGRSPDGRTHRNICKYGADLYVDRPSPDLAGYSLAPYSRAAAPIDPG----- 459
      |      |      |      |      |      |      |      |      |      |
Db 465 CSSGGGAVNRGALGPPGL-LVLAAG--DRRGEQTAVFYVSRGLDGLHTSFQDELRRS 521
      |      |      |      |      |      |      |      |      |      |
Qy 460 -----AKSVHLRIILVDTSQSVFVFNAGHTVLSSQVHPAEG--DTGIS 499
      |      |      |      |      |      |      |      |      |      |
Db 522 RAKDVTNRVIGSTVPVLDGEGALSMRVLVDHSIVOGEDMGGRTTTMRVYPMESYQEARVY 581
      |      |      |      |      |      |      |      |      |      |
Qy 500 LYTDGGAHFTG--IVVREIGQA 520
      |      |      |      |      |      |      |      |      |      |
Db 582 LFNNATGASVTAERLVVHEMDSA 604
      |      |      |      |      |      |      |      |      |      |

RESULT 11
US-10-003-392-20
; Sequence 20, Application US/10003392
; Patent No. US20020170086A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Calmi, Perry G.
; APPLICANT: Stoop, Johan M.
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BB1463 US NA
; CURRENT APPLICATION NUMBER: US/10/003,392
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-10-003-392-20

```

Query Match	5.6%	Score 160;	DB 9;	Length 600;
Best Local Similarity	21.2%;	Pred. No. 4.6e-05;		
Matches	126;	Conservative 78;	Mismatches 244;	Indels 146;
Gaps				
Qy	24	GGAVPPAARASAFGS-----LRAYHMTPPSGWGLCDQRPQPVTHHAGYQLYLYLH-----SD 73		
Db	42	GLRYDEEAAAGFPNSNEMLQWQSGYHFQTAKNYMSDPNGLMYNGWYHMFQYNPVGTD 101		
Qy	74	QNNPGCGGDHASTTGDGVAFTHHGTVMPLRDPFVW-----SGSAVV-----GTA-----NTA 120		
		: :		
Db	102	WDDG--EWGHAVSRNLVW-----RTLPIAMVADQWDLGLVLSGSMVPLPNGTVIMYYTG 156		
Qy	121	GFGAGAV-----VALATOPTDGV-----RKYEQYLYWSTDGGFTTALPDFVIVNTDGRAA 172		
Db	157	ATNASAVEVOCIATPADPNDFELRRWTKHPANPVIWSPPG-----IGTKD-----201		
Qy	173	TPPAEINAEWFRDPKTHW-DTARGEWCVCIGRL-----RYAIFYTSPNLRDWTLRR 223		
Db	202	-----FRDPTAWYDESDDTWRTLGLSKDDHGDHGDIAIMMYKTQFNLNYELIP 250		
Qy	224	NFDYPNHALGIECPOLFEI--PADDGTTHW--VLAASMDAVGIGLPMYAYWT--GTWDG 278		
		: :		
Db	251	GIILRVORTGEWICIDPYGVGRNSNDSMLHLVKASMD-----DERHDYISLGTYS 304		
Qy	279	EQFHADDLTPQWLDWG-----WDW--YAAVTPSIDAPETKRLAIAMNNWYAAARDVPT 331		
Db	305	AANAWTPIDPE-LDLGILRYDNGKPYASTSF--YDPAKKRRVLMGYVEGVDSKRAADV--359		
Qy	332	DASDYGQNSIVREILRLARQPGGWYTLTSTPVAALTNVYVTAFTTLPLDRTVDGSAVLP--389		
		: :		
Db	360	--VKGWASIQSVPRTIALDEKTR--TNLLLPWVEETLRLNATELSDVTLNTGSIHIP 415		
Qy	390	WNGRAYEIE-----LDIAWDATN-----VGISVGRSPDGTHTNNIGYKGLADLYVDRGPSDL 441		
Db	416	LROGTQLDIEATFHLDASAVALNEADVNGYSSSGGAVNRGALGPFGL-LVLAAG--DR 472		

QY 442 AGYSLAPYSRAAPIDPG-----ARSHLRILVD 470
Db 473 RGEQTAIVFYVSRGLDGLHTSFCDELSSRAKDVTKRVIGSTVPVLDDGEAFSMRVLVD 532
QY 471 TQSVVEFVNAGHTVLSQQVHFAEGDTGISLY-----TDGPAHTGIVVREIGQA 520
Db 533 HSIQGFAMGGRTTMTSRYPMEAYOEAKVYLFNATGASVMAERLVVHEMDSA 586
RESULT 12
US-10-003-392-4
; Sequence 4, Application US/10003392
; Patent No. US20020170086A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Caimi, Perry G.
; APPLICANT: Stoop, Johan M.
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BB1463 US NA
; CURRENT APPLICATION NUMBER: US/10/003,392
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 609
; TYPE: PRT
; ORGANISM: Parthenium argentatum Grey
US-10-003-392-4
Query Match 5.4%; Score 155; DB 9; Length 609;
Best Local Similarity 20.9%; Pred. No. 0.00012;
Matches 117; Conservative 67; Mismatches 193; Indels 184; Gaps 31;
QY 40 RAVYHMTPPSGWLCDPORPVTHGAYQLYLHSDQNN--GPGGWDHASTTDGVAETHHGT 97
Db 81 RTAFHFQPAKNIYDPNGQMYMGVHLYFYQNPYAPVWGNMWSGHSVSKDMINMY---- 136
QY 98 VMLPRDFPWSGSVAVGTA--NTAGFAGAGAVVALATQTDGVRKYOEQLYVWSDGGTT 155
Db 137 -----ELFP-----AIVPTWYDIEGVLSGTSIVLPG-----QIFALYTGANDFS 178
QY 156 ---FTALP-----DPVIVN-----TDGRAATTAEIENAEWFRDPKIHWDTARGEWCVICR 204
Db 179 QLOKRAVPNSDPLLVWVKVEDNPILYTPPGI--GLKYRDPSTVWGPDPGKHRMIMGT 237
QY 205 LR-----YAAFYTSPNLRDWTLRRNFDYPNHALGGI---ECPDLFEITADD----- 247
Db 238 KRNGTGMILVYHTT---DYTNEMLENEPMHVSVPNTDMWECVDFYPVSLTNSALDIAAY 293
QY 248 --GTRHWVLAASMDAYGIGLPMYAYTGTWGDGEFHADDLTPQ--WLDWG-----WDWVA 299
Db 294 GSGIRK-VIKESWEGYGMDF-----YSICTYDA---FNDKWTPDPNPELDVGIGLRCDYGR 344
QY 300 AVTWPSIDAPETKRLAIAMNMNWKYAADVPDTPDASDGYNGQNSIVRELRLARQPGGWYTL 359
Db 345 PFASKSIEDPVKKRI-----TWAVYGE-----SD--NADDLSR-----GWATV 382
QY 360 LSTPVAALTNVYATTTLPDRVDCGSVLPW-----NGRAY-EIELD-----IA 402
Db 383 -----YNVGRTIVLDKRT-GTHLLHWPEVIESLRYNGQFEKEIKLPGSIAPLD 431
QY 403 WDTATNVGI-----SVGRSPDGRTHNICKYGADLYVDRGPSDL 441
Db 432 IGTATQLDIAFFKVDAAALNATSETDNNFACTTSSGAVERGSIGLPGFLAVLADGTLSEL 491
QY 442 AGYSLAPYSRAAPIDPGARSVH-----LRLIV 469
Db 492 T-----PVYFYTAKKADGGV-STHFCTDKLRSLDFDKERVVYGVSTVPVLDDDEELTMRLIV 546

QY 470 DTQSVVEFVNAGHTVLSQQVH 490
Db 547 DHSVVEAFQAQGGRIATSRVY 567
RESULT 13
US-09-990-385-11
; Sequence 11, Application US/09990385
; Publication No. US20020192771A1
; GENERAL INFORMATION:
; APPLICANT: Koji YANAI et al.
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF
; ISOLATING BETA-FRUCTOFURANOSIDASE GENE, SYSTEM FOR PROD
; BETA-FRUCTOFURANOSIDASE, AND BETA-FRUCTOFURANOSIDASE V
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., Suite 800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/990,385
FILING DATE: 10-Sep-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/142,623
FILING DATE: September 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: 2001-1611
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acid residues
TYPE: Amino acid
STRANDEDNESS: No. US20020192771A1 relevant
TOPOLOGY: No. US20020192771A1 relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Microorganism: Penicillium roqueforti IAM7254
FEATURE:
NAME/KEY: mat peptide
LOCATION: 1..565
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-990-385-11
Query Match 5.2%; Score 147.5; DB 9; Length 565;
Best Local Similarity 21.8%; Pred. No. 0.00046;
Matches 126; Conservative 68; Mismatches 226; Indels 159; Gaps 29;
QY 26 AVPPAARASAPGSL-----RAVYHMTPPSGWLCDP-----QRPVTHGAYQLYLHSDQNN 77
Db 11 SAPPNLSTLANASLFKTRPRAHLPPSGNIGDPCGHYTDPKT--GLFHWGLYS-----G 64
QY 78 PGGDWHAISTTDGVAFTHHGTVMLPRDPFPVWGSASVVGTA--TAGFGAGAVV--ALATOP 134
Db 65 ISG-----ATTDDLIV-----TYKDLNPD---GAPSIVAGGKNDPLSVFDGSIPIGDMGP 112
QY 135 T-----DGVRYKOEQLYVWSDGGTFTALPDPVIVNTGRAATTPAEIE 179
Db 113 TLLYTSVSVLPINHWSIPYTRGSETQSLAVSYDGGHNFTKL-----NOGPVITPPFAL 165
QY 180 NAEWFRDPKIHW-----DTARGEWCVICG-----RLRYAAFYTSFNLRL 218

Db 166 NVTAFRDPYVFQSPILDKSVNSTQGTWYVAISGGVHGVCQFLYRQNDADFQYWEYLQ 225
QY 219 W---TLRRNFDPYVFNHALG---GTECPDLPEITADGTRHVVLAASMDAYGIGLP----- 266
Db 226 WKEPLNTTWCKGDWAGGWFNEVGNVSLNAGVSEDEGEIITITGAEGLPIVPOVS 285
QY 267 ---MTYAYWTGMDGEQFHADDLTQWLDGMDWYAAV---TWPSIDAPETKR-----L 314
Db 286 SIRDMLVNTGNVNDGVSFTKPTMAGV-LDWGVSAYAAAGKILPASSQASTKSGAPDREI 344
QY 315 AIAMNNWKY-AARDVPTDASDYNQNSIVREL-----RLARQ 352
Db 345 SYVWLTGDLFEQVKGFPPT-AQQNWTGALLPRELNVRTISNVVDNELSRESLTSWRVARE 403
QY 353 PGGWYLLTSPVAALTNVYATTT---LPDRVTGSAVLPWN---GRAYEIELDIW 403
Db 404 DSCQIDLEWGISISRETVSALTSFGSFSVSGKTLNAGAVPNTSPSSKFFVLTANISF 463
QY 404 DT-ATNVGISVGRSPDGRHTNIGKY---GADLYYDRGPSDLAGYSLAPYSRAAAPI-- 456
Db 464 PTSARDSGIQAGFOVLSSSLESTIYQFSNESIIVDRSNTSAA-----ARTTAGILS 516
QY 457 ---DPG-----ARSHLRLVDTQSVFVN 479
Db 517 DNEAGRRLFDVLRNGKEQVETLELTIVVDNSVLEYIAN 555

RESULT 14

US-09-990-385-1
; Sequence 1, Application US/09990385
; Publication No. US20020192771A1
; GENERAL INFORMATION:
; APPLICANT: KOJI YANAI et al.
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF
; ISOLATING BETA-FRUCTOFURANOSIDASE GENE, SYSTEM FOR PRODUCTION
; OF BETA-FRUCTOFURANOSIDASE, AND BETA-FRUCTOFURANOSIDASE VARIANTS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/990,385
; FILING DATE: 10-Sep-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/142,623
; FILING DATE: September 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: 2001-1611
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 amino acid residues
; TYPE: amino acid
; STRANDEDNESS: No. US20020192771A1 relevant
; TOPOLOGY: No. US20020192771A1 relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Microorganism: Aspergillus niger ACE-2-1
; (ATCC 20611)

FEATURE:
NAME/KEY: mat peptide
LOCATION: 1..635
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-990-385-1

Query Match 5.0%; Score 141.5; DB 9; Length 635;
Best Local Similarity 19.1%; Pred. No. 0.0017;
Matches 125; Conservative 83; Mismatches 245; Indels 203; Gaps 28;

QY 26 AVPPAARASAPGSL-----RAVYHMTPPSWLWCDPQRPVT--THGAYQLYLLISDQNGNP 78
Db 8 APPPTLSTLPNTLTFHVPRAHILPAEQIGDPCAHYTDPTSTGLFHVGLH--DGDGI 65
QY 79 GGDWAS-----TTDGVAFTHHGTVMPLRPDPF---VMSGSAV-VGTANTAGCAGAVWA 129
Db 66 AGATTANLATYTTDSNGSF---LIQPGKNDPVAVDGAVIPGVNNTPTLLYTSVSF 121
QY 130 LATQ-PTDGVYRKQEOYLYWSTDGFTF-----TALPDP----- 162
Db 122 LPIHWSIPYTRGSETQSLAVARDGGRREFDKLDQGPVIADHPFAVDVTAERDPFVFRSAKL 181
QY 163 -VIVTDGRAATTPAEIEA--EWRDPKIHWDTARGWCVCIGRLRYAAFTSPNLRD-- 218
Db 182 DVLLSLDEEVARNETAQQAVDGVGTEKNAPYVAVSGVGHVGPQAQFLYRQNGGNASEFQ 241
QY 219 -----WTLRRNFDYPNHALGGI-----ECPDLFEITADDT 249
Db 242 WYELGEMWQEQATNSSWGDGTWAGWGF---NFETGNVLFLEEGHDPQGTGEVFTLGT 298
QY 250 RHWVLAASMDAYGIGLPMT-----YAYWTGTWGDGEQFHADDLTPQ---WLDWGW 295
Db 299 E-----CSGLPIVPOVSSIHDMLWAAGEVGVSGEAGAKVEFSPSMAGFLDWGE 347
QY 296 DWYAAV--TWPSIDABE-----TKRLATAMNNWKYAAARDVPTDASDYNQNSIVRE 346
Db 348 SAYAAAAGKVLPAASAVSKTSVGVVDVRYVFWLTGQYEQADGFPPTAQQGWTSLLPRE 407
QY 347 LR-----LARQPG-----GWYTLSTPVAALTNVYATTT 377
Db 408 LKVQTVENVVDNELVREEGSVWVGESDQNTARLTGITIARETKAALLAN--GSVTAE 465
QY 378 PDRTVGSAVLPW---NGRAYEIELDIAWDATYN-----VGISVGRSPDGRHT 423
Db 466 EDRTLQATAAVVPAQSPSSKFFVLTAQLEPPASARSPLQSGFEILASELERTAIYYQFS 525
QY 424 NICKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPG-----AR 461
Db 526 N-----ESLVVDRSQT-----AAAPNGLDSFTESGKRLFDVIENGQEQVE 569
QY 462 SVHLRLVDTQSVFVFNAGHTVLSQOVHFAEGDGTGISLYTDG-GPAHFTGIYVRE 516
Db 570 TLDLTVVVDNAVVEYANGFRFALSTWARSWYDNTQIRFFHNGEVEGVQFRNVSVSE 625

RESULT 15

US-09-990-385-13
; Sequence 13, Application US/09990385
; Publication No. US20020192771A1
; GENERAL INFORMATION:
; APPLICANT: KOJI YANAI et al.
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF
; ISOLATING BETA-FRUCTOFURANOSIDASE GENE, SYSTEM FOR PRODUCTION
; OF BETA-FRUCTOFURANOSIDASE, AND BETA-FRUCTOFURANOSIDASE VARIANTS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/990,385
FILING DATE: 10-Sep-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/142,623
FILING DATE: September 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: 2001-1611
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 574 amino acid residues
TYPE: Amino acid
STRANDEDNESS: No. US20020192771A1 relevant
TOPOLOGY: No. US20020192771A1 relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Microorganism: Scopulariopsis brevicaulis IFO4843
FEATURE:
NAME/KEY: mat peptide
LOCATION: 1..574
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-990-385-13

Query Match 4.9%; Score 140.5; DB 9; Length 574;
Best Local Similarity 21.9%; Pred. No. 0.0018;
Matches 129; Conservative 56; Mismatches 245; Indels 159; Gaps 29;

QY 26 AVPPAARASAPGSL----RAVYHMPPPSGWLCDPQRPVT--THGAYQLYLHSDQNNPG 79
DB 19 SAPPNLSTLANNLETPRAHVLPPQNOIGDCPMHYTDPETGIFHGWLYN--GNGAS 76
QY 80 GWDHASTDGVAFTHGTVMLRPDPFVWMSGAVV---GTANTAGFGAGVV--ALATOP 134
DB 77 G---ATTEDLV-----TYQDLNPD-----GAQMILPGGVNDPLAVFDGAVIPSGIDGKP 122
QY 135 TDGVRKYQEQLYV-----STDGGFTFTAL-PDPVIVNTDGRAATP---- 175
DB 123 TMNMTSVSYMPISWISYATRGSETHSLAVSDGGKNTKLVQGPVTPSPFPGANVTSWRD 182
QY 176 -----AEIENAEWF-----RDPKIHWDTARGEWVCVIGR 204
DB 183 PFLFQNPQFDSLESENGTWYTVISGGIHGDGPSAFLYRQHDPDFQYWEYLGFWNNEGN 242
QY 205 LRYAIFYTSPNLRDWTLRNDYPNHALGGIE----CPDLFEITADDGTRHWVLAASMDA 260
DB 243 STWGS-----GDWAGRWGNFVINIVGLDDGDYNDP-GEIFATVGT-EWSFDPIKPKQ 293
QY 261 YGIGLPMTYAYWTGTWGDGEQFHADDLTPOWLDGWDAVAAV--TWPSIDAPETKR----- 313
DB 294 ASDNREMLWAAGNMTLEGDGIKFTPSMAGYLDWGLSAYAAAGKELPASSKPSQKSGAPDR 353
QY 314 -LAIAMNNWYKARDVPTDASDYGNGQNSIVREL-----RLAROPGGWYTLIS 361
DB 354 FVSYLWLTGDFEGHDFPT-POONTGSLLLPRELSVGTIPNVVDNELARETGSW--RVG 410
QY 362 TPVAALTNVVT-----ATTLPDRTVGDS-----AVLPWNGRAYEIELDIAMDATNVG 410
DB 411 TNDTCVLELVTLKQEIARETAEMTSGNSFTESARNVSSPGSTAFQOSLDSKFFVLT-AS 469
QY 411 ISVGRSPDGTTRTNICKYGADLYVDRGPDSDLAGYSLA-----PYSRAAPIDPG---- 459
DB 470 LSF---PSSARDSL-KAGFEILSEFEFTVYQFNSNIIIDRSNSAAALTTDGDIT 525

Search completed: May 6, 2003, 12:44:10
Job time : 21.5337 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:32:47 ; Search time 58.6125 Seconds
(without alignments)
1515.790 Million cell updates/sec

Title: US-09-868-328B-1
Perfect score: 2845
Sequence: 1 MTPAISRRVLOGAGAGALA.....TDGPAHFTGIVVREIQAI.521

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 785308.seqs, 170526417 residues

Total number of hits satisfying chosen parameters: 785308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2845	100.0	521	5	US-09-868-328A-2
2	2845	100.0	521	5	US-09-868-328B-1
3	2837	99.7	519	5	US-09-868-328B-4
4	944	33.2	516	6	US-10-369-493-23310
5	556	19.5	677	6	US-10-369-493-23212
6	449.5	15.8	541	6	US-10-369-493-13045
7	359.5	12.6	532	6	US-10-369-493-22074
8	348	12.2	581	6	US-10-369-493-22749
9	313.5	11.0	311	6	US-10-369-493-10616
10	304	10.7	670	1	PCT-US03-07858-1502
11	304	10.7	670	6	US-10-289-757-176
12	304	10.7	670	6	US-10-389-566-1502
13	295	10.4	490	6	US-10-369-493-3411
14	294	10.3	668	6	US-10-289-757-87
15	291	10.2	677	1	PCT-US03-07858-1977
16	291	10.2	677	6	US-10-389-566-1977
17	285.5	10.0	583	6	US-10-289-757-83
18	283.5	10.0	583	6	US-10-289-757-79
19	282	9.9	662	1	PCT-US03-07858-789
20	282	9.9	662	6	US-10-389-566-789
21	279	9.8	636	1	PCT-US03-07858-1296
22	279	9.8	636	6	US-10-389-566-1296
23	279	9.8	646	1	PCT-US03-07858-1505
24	279	9.8	646	6	US-10-389-566-1505
25	268.5	9.4	673	1	PCT-US03-07858-482
26	268.5	9.4	673	6	US-10-389-566-482

ALIGNMENTS

RESULT 1

US-09-868-328A-2
; Sequence 2, Application US/09868328A

; GENERAL INFORMATION:

; APPLICANT: RHEE, Sangki

; APPLICANT: SONG, Kibang

; APPLICANT: KIM, Chulho

; APPLICANT: RYU, Eunja

; APPLICANT: LEE, Yongbok

; TITLE OF INVENTION: Enzymatic Production of Diffructose Dianhydride IV from sucrose
; TITLE OF INVENTION: relevant enzymes and genes coding for them

; FILE REFERENCE: 24679

; CURRENT APPLICATION NUMBER: US/09/868,328A

; PRIOR FILING DATE: 2001-06-18

; PRIOR APPLICATION NUMBER: PCT/KR00/01183

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 521

; TYPE: PRT

; ORGANISM: Arthrobacter ureafaciens

US-09-868-328A-2

Query Match

Best Local Similarity 100.0%; Score 2845; DB 5; Length 521;

Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTPAISRRVLOGAGAGALIFGGAVPPAARASAPGSLRAVYHMTPPSGWLCDDPORPVT 60

Db 1 MTPAISRRVLOGAGAGALIFGGAVPPAARASAPGSLRAVYHMTPPSGWLCDDPORPVT 60

QY 61 THGAYQLYLHSDNNGPGWDBHASTTDCGVAFTHGCTVMPLRPFDPFVWSSGSAVVGTA 120

Db 61 THGAYQLYLHSDNNGPGWDBHASTTDCGVAFTHGCTVMPLRPFDPFVWSSGSAVVGTA 120

QY 121 GFGAGAVVALATQPTDGVKRYOEQYLYWSTDGGFTTALPDVIVNTDGRAATTPAEIN 180

Db 121 GFGAGAVVALATQPTDGVKRYOEQYLYWSTDGGFTTALPDVIVNTDGRAATTPAEIN 180

QY 181 AEWFRDPKTHWDTARGEWCVCIGRLRYAFTYSPNLRDWTLRNFDPYPHNALGGIECPDL 240

Db 181 AEWFRDPKTHWDTARGEWCVCIGRLRYAFTYSPNLRDWTLRNFDPYPHNALGGIECPDL 240

QY 241 FEITADGGRHWVLAASMDAYGIGLPMYAYWTGTWDGQGFHADDLTLPWLWDGWDWYAA 300

Db 241 FEITADGGRHWVLAASMDAYGIGLPMYAYWTGTWDGQGFHADDLTLPWLWDGWDWYAA 300

QY 301 VTWPSIDAPETKRLAIAWNNMKYAAARDVPTDASDGYNGQNSIVRELRLARQPGWYTL 360

Db 301 VTWPSIDAPETKRLAIAWNNMKYAAARDVPTDASDGYNGQNSIVRELRLARQPGWYTL 360

Db 301 VTWPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRLARPGGWYTL 360
QY 361 STPVAALTNVYVATTTLDPRTVDGSAVLPNWGRAYEIELDIAMDATNVGISVGRSPDGT 420
Db 361 STPVAALTNVYVATTTLDPRTVDGSAVLPNWGRAYEIELDIAMDATNVGISVGRSPDGT 420
QY 421 RHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVEFVNA 480
Db 421 RHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVEFVNA 480
QY 481 GHTVLSQQVHFAEGDTGISLYTDGGPAHFTGIVVREIGQAI 521
Db 481 GHTVLSQQVHFAEGDTGISLYTDGGPAHFTGIVVREIGQAI 521

RESULT 2

US-09-868-328b-1

; Sequence 1, Application US/09868328B

; GENERAL INFORMATION:

; APPLICANT: RHEE, Sang-Ki

; APPLICANT: SONG, Ki-Bang

; APPLICANT: KIM, Chul-Ho

; APPLICANT: RYU, Eun-Ja

; APPLICANT: LEE, Yong-Bok

; TITLE OF INVENTION: Enzymatic production of difructose dianhydride IV from sucrose and relevant enzymes and genes coding for them

; FILE REFERENCE: 24679

; CURRENT APPLICATION NUMBER: US/09/868,328B

; CURRENT FILING DATE: 2001-06-18

; PRIOR APPLICATION NUMBER: PCT/KR00/01183

; PRIOR FILING DATE: 2000-10-19

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 1

; LENGTH: 521

; TYPE: PRT

; ORGANISM: Arthrobacter ureafaciens K2032

US-09-868-328b-1

Query Match 100.0%; Score 2845; DB 5; Length 521;
Best Local Similarity 100.0%; Pred. No. 2.7e-209; Mismatches 0; Indels 0; Gaps 0;
Matches 521; Conservative 0;

QY 1 MTPAISRRVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDPQRPVT 60
Db 1 MTPAISRRVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDPQRPVT 60
QY 61 THGAYQLYLYLHSDQNGPCGWDHASTTDGVAFTHHGTVMPLRDPDPVWGSASVVGVTANTA 120
Db 61 THGAYQLYLYLHSDQNGPCGWDHASTTDGVAFTHHGTVMPLRDPDPVWGSASVVGVTANTA 120
QY 121 GFGAGAVVALATOPTDGVVRKYQEOYLWSTDGGFTTALPDPIVNTDGRAATTAEIEN 180
Db 121 GFGAGAVVALATOPTDGVVRKYQEOYLWSTDGGFTTALPDPIVNTDGRAATTAEIEN 180
QY 181 AEWFRDPKTHWDTARGEWCYVIGRLRYAAYTSPNLRDWTLRNFDYPNHALGGIECPDL 240
Db 181 AEWFRDPKTHWDTARGEWCYVIGRLRYAAYTSPNLRDWTLRNFDYPNHALGGIECPDL 240
QY 241 FEITADDGTRHWVLAASMDAYGIGLPMTYAYWTGTWDSQFHADDLTPQWLWDGWDWYAA 300
Db 241 FEITADDGTRHWVLAASMDAYGIGLPMTYAYWTGTWDSQFHADDLTPQWLWDGWDWYAA 300
QY 301 VTWPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRLARPGGWYTL 360
Db 301 VTWPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRLARPGGWYTL 360
QY 361 STPVAALTNVYVATTTLDPRTVDGSAVLPNWGRAYEIELDIAMDATNVGISVGRSPDGT 420
Db 361 STPVAALTNVYVATTTLDPRTVDGSAVLPNWGRAYEIELDIAMDATNVGISVGRSPDGT 420
QY 421 RHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVEFVNA 480

Db 421 RHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVEFVNA 480
QY 481 GHTVLSQQVHFAEGDTGISLYTDGGPAHFTGIVVREIGQAI 521
Db 481 GHTVLSQQVHFAEGDTGISLYTDGGPAHFTGIVVREIGQAI 521

RESULT 3

US-09-868-328b-4

; Sequence 4, Application US/09868328B

; GENERAL INFORMATION:

; APPLICANT: RHEE, Sang-Ki

; APPLICANT: SONG, Ki-Bang

; APPLICANT: KIM, Chul-Ho

; APPLICANT: RYU, Eun-Ja

; APPLICANT: LEE, Yong-Bok

; TITLE OF INVENTION: Enzymatic production of difructose dianhydride IV from sucrose and relevant enzymes and genes coding for them

; FILE REFERENCE: 24679

; CURRENT APPLICATION NUMBER: US/09/868,328B

; CURRENT FILING DATE: 2001-06-18

; PRIOR APPLICATION NUMBER: PCT/KR00/01183

; PRIOR FILING DATE: 2000-10-19

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 4

; LENGTH: 519

; TYPE: PRT

; ORGANISM: Arthrobacter ureafaciens K2032

US-09-868-328b-4

Query Match 99.7%; Score 2837; DB 5; Length 519;
Best Local Similarity 100.0%; Pred. No. 1.1e-208; Mismatches 0; Indels 0; Gaps 0;
Matches 519; Conservative 0;

QY 1 MTPAISRRVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDPQRPVT 60
Db 1 MTPAISRRVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDPQRPVT 60
QY 61 THGAYQLYLYLHSDQNGPCGWDHASTTDGVAFTHHGTVMPLRDPDPVWGSASVVGVTANTA 120
Db 61 THGAYQLYLYLHSDQNGPCGWDHASTTDGVAFTHHGTVMPLRDPDPVWGSASVVGVTANTA 120
QY 121 GFGAGAVVALATOPTDGVVRKYQEOYLWSTDGGFTTALPDPIVNTDGRAATTAEIEN 180
Db 121 GFGAGAVVALATOPTDGVVRKYQEOYLWSTDGGFTTALPDPIVNTDGRAATTAEIEN 180
QY 181 AEWFRDPKTHWDTARGEWCYVIGRLRYAAYTSPNLRDWTLRNFDYPNHALGGIECPDL 240
Db 181 AEWFRDPKTHWDTARGEWCYVIGRLRYAAYTSPNLRDWTLRNFDYPNHALGGIECPDL 240
QY 241 FEITADDGTRHWVLAASMDAYGIGLPMTYAYWTGTWDSQFHADDLTPQWLWDGWDWYAA 300
Db 241 FEITADDGTRHWVLAASMDAYGIGLPMTYAYWTGTWDSQFHADDLTPQWLWDGWDWYAA 300
QY 301 VTWPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRLARPGGWYTL 360
Db 301 VTWPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRLARPGGWYTL 360
QY 361 STPVAALTNVYVATTTLDPRTVDGSAVLPNWGRAYEIELDIAMDATNVGISVGRSPDGT 420
Db 361 STPVAALTNVYVATTTLDPRTVDGSAVLPNWGRAYEIELDIAMDATNVGISVGRSPDGT 420
QY 421 RHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVEFVNA 480
Db 421 RHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVEFVNA 480
QY 481 GHTVLSQQVHFAEGDTGISLYTDGGPAHFTGIVVREIGQ 519
Db 481 GHTVLSQQVHFAEGDTGISLYTDGGPAHFTGIVVREIGQ 519

RESULT 4

US-10-369-493-23310
; Sequence 23310, Application US/10369493
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 23310

; LENGTH: 516
; TYPE: PRT

; ORGANISM: Bacillus subtilis
US-10-369-493-23310

Query Match 33.2%; Score 944; DB 6; Length 516;

Best Local Similarity 41.8%; Pred. No. 8.6e-64;

Matches 205; Conservative 71; Mismatches 176; Indels 38; Gaps 15;

QY 40 RAYHMTPPSGWLCDDPQRPVTHGAYQLYLLHS--DONNGPG--GWDHASTTDGVAETHHGT 97

DB 44 RAYHMTTPDKKNDQKPIYDGKHYLYNRDPKNGTEWTHAVSEDLVHWTDEGV 103

QY 98 VMP--LRPDPFVWSSGAVVGTANTAGGAGAVVALATOPTDGVKRYQEOYLWSTDGGFT 155

DB 104 AIPKVTNPDGDIWTSVVVDKENTAGFGKNALVAIVTQPSAKDKK--QEOYLWSTDKGS 162

QY 156 F-----TALPDVIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGEWCWVIGRLRYAA 209

DB 163 FRYSGNPMNP--GTDD-----FRDPKVIWDDQDNKVMVMAEGSKTG 205

QY 210 FYTSNLRDWTLLRNFDYPNHALGGTECPDLFEITADGTRHWLAASMDAYGIGLPMY 269

DB 206 FYESNLKDNHYSYSGF--FPEQA--GMVECPDLYMMRASDGTNKLWLGASANGPKWPKNTY 263

QY 270 AYWTGTWGEQFADDLTPQLWDGWDWYAAVTPSIDA--PETKRLATAWNNNNKYAAR 327

DB 264 AYWTSFDCGKFKADQTEAQWLDYGFWDYGGVTFEEDSKSTDPLEKRYALAWNNNDY--AN 322

QY 328 DVPTDASDGYNGQNSIVRELRLAROPGQWYTLSTPVAALTNVYVATTPLPORTVDGSAY 387

DB 323 NTPT--MKNFGNCTSVIRELRKEQ--DGTYSLSVQPIEALEQLTVSTDEIQDQVNGSKT 380

QY 388 LPWNGRAYEILDIAWDATNNGISVGRSPDGTNRHNTICKY--GADLYVDRGPSDLAGYS 445

DB 381 LSIITGDTOLDLWSELKNAVRLESEDKRHLIDVGFAGGAYVYVNRATNPDKS 440

QY 446 LAPYRAAAPIDPGARSVHLRLIVDTQSVVEFVNAAGHTVLSQQVHFAEGDTGSLVTDGG 505

DB 441 -NTYESKAPYDVNKKRKHILVDKTTIEVFVGDKTVFSNEVPKPKEDKGTILYSOGG 499

QY 506 PAHFTGIVVR 515

DB 500 TASFRNITVK 509

RESULT 5

US-10-369-493-23212

; Sequence 23212, Application US/10369493
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 23212

; LENGTH: 677
; TYPE: PRT

; ORGANISM: Bacillus subtilis
US-10-369-493-23212

Query Match 19.5%; Score 556; DB 6; Length 677;

Best Local Similarity 32.2%; Pred. No. 6.3e-34;

Matches 160; Conservative 74; Mismatches 217; Indels 46; Gaps 18;

QY 40 RAYHMTPPSGWLCDDPQRPVTHGAYQLYLLHS--DONNGPG--GWDHASTTDGVAETHHGT 97

DB 35 RPQYHFTPEANWMDPDMGVVYAGEYHLYFYQHPYGLQWPMHGHAVSKDLVTWEH--L 92

QY 98 VMPLRPD--FPVWSSGAVVGTANTAGFGAG---AVVALATOPTDGVKRYQEOYLWSTDG 152

DB 93 PVALYPDEKGTIFSGSAVVDKXNNTSGFTGKEKPLVAIYTDREG---HVOVSIAYSNDK 149

QY 153 GFTFTALP--DPVIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGEWCWVIGRLRYAAFY 211

DB 150 GRITWKYAGNPVLPN-----PGKDK---FRDPKVFWEYKEKKWVWVLAAGDRILLY 197

QY 212 TSNPLRDLTLLRNFDYPNHALGGI--ECPDLFEITAD--DGTNRHVLAAASMDAYGIGLPMY 268

DB 198 TSKNLKQWYASEFGDQSGHGVWECPLDFELPVDGNPNQKKWVWVSGVNGAVSGSG 257

QY 269 YAYWTGTWGEQFADDLTPQ--WLDWGDWYAAVTPSIDAPETKRLATAWNNNNKYAA 326

DB 258 MQYFVGDFDGTGTHFKNENPNKVLWTDYGRDYAAVSWSDIPSTDSRRLWLGWMSNQY--A 316

QY 327 RVPTDASDGYNGQNSIVRELRLAROPGQWYTLSTPVAALTNVYVATTPLPORTVDGSA 386

DB 317 NDVPTSP---WRSAYTSIPRELKKAFTGEG--VRVQTPVKELETIRGTSKKNNLITSPAS 372

QY 387 ---VLPWNGRAYEY---ELDIAMDITATNNGISVGRSPDGTNRHNTICKYGADLYVDRGPSDL 441

DB 373 HNVLAGQSGDAYEINAEFKVSPGSAAEFGFKVKTGENQFTKVDYDRNNAKLFVDRSECN 432

QY 442 AGYSLA--PYSRAAAPIDPGARSVHLRLIVDTQSVVEFVNAAGHTVLSQQVHFAEGDTGSL 500

DB 433 DTENPAFNTGKETAPLKPVNGKVKLRFYVDRSSVEFVNGDKQVITDIILPDRSSKGLEL 492

QY 501 YTDGGFAHFTGIWVREI 517

DB 493 YAANG-----GVKVKSL 504

RESULT 6

US-10-369-493-13045

; Sequence 13045, Application US/10369493
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13045

; LENGTH: 541
; TYPE: PR
; ORGANISM: Aspergillus nidulans
US-10-369-493-13045

Query Match 15.8%; Score 449.5; DB 6; Length 541;
Best Local Similarity 28.1%; Pred. No. 6.5e-26;
Matches 160; Conservative 77; Mismatches 205; Indels 127; Gaps 29;
Db 40 RAVYHMTTPSGWLCDDPQRTTHGAYLYLHSDONNGPGG-----WDHASTTD--- 88
Db 6 RPQYHFTPAQWINDPNGLYHNGTYHMFYQYN-----PGVAHEAMSWGHATSQDILTHW 60
QY 89 ---GVAETHGTVMPLRDPFP-----VWSSAVVGTANTAGFG-----AGAVVALAT- 132
Db 61 DERPVALTARG-----FPGNISQFFSGSAVDVNDSSGNGMGTAAPWAMVYTS 112
QY 133 -----QPTDGVK-----YOEQYLWSTGGFTFTA--L 159
Db 113 FVCYPSLLIEWSDLEMLVLTCTVCLGQTLPSGKKVRASQOSQSLAYSLDSGETWTDDA 172
QY 160 PDVIVNTDGRAATTPAEIENAEWFRPKIHWDTARGEW--CVIGRLRYAAFTSPNLR 217
Db 173 VNPILQ-----PPHPYESQYKEM-RDPNVFVHAPYKWIITLALSOLHLLIYTSDLK 226
QY 218 DWTLRNFDYPNHALGGI-ECPLFEITADDGTRH--WVLAASMDAYGIGLPMTYA--- 270
Db 227 NWTLASEFG-PYNAVGVWBCPSFFPLPVDNDESNNKVAIIGLNP--GPPCTVSGNQ 283
QY 271 YWTGTWDEGF--HADDLTQ-----WLDGWDMYAAVTFWSDAPETKRLAIAMNNKY 324
Db 284 YILQFNGTHFLPDAESLHDEGEANWLDYDGFYALVYNGL--PEQRTVIAMNSMNOY 341
QY 325 AARDVPTDASDGYNGQNSIVRELRL-----AROP-GGWYTLTSTPVAALTNVYAT 374
Db 342 AEK-VPTV---WRNANTIPRLCLTKINGKVLVQPEANWAAIMSKP-----QTTRI 391
QY 375 TTLPDRTVDSGAVLPWNGR--AVEIELDIAMDATNNGVSGRSPDGTNRHNGKYGA-- 430
Db 392 LSLPANSTRNLGSL--GKTLAANLTFSAEQNGSGSFGISLLATQNTTQTLIGYNFATK 448
QY 431 DLVYDRGPSLAGYSLAPYSRAAPIDPG--ARSVHLRLVDTOSVEVFNAGHTVLSQOV 489
Db 449 HIFVDRASGNTSDETFASVYRAPLEPADDDGTTLRIFADMSVEVFGCGGETTLTAQV 508
QY 490 HFAEGDTGISLYTDGGFPAHFTGIWVREIG 518
Db 509 FPEGATHARLFANGAGVEDVLLSVSKVG 537

RESULT 7
US-10-369-493-22074
; Sequence 22074, Application US/10369493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO: 22074
; LENGTH: 532
; TYPE: PR
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22074

Query Match 12.6%; Score 359.5; DB 6; Length 532;
Best Local Similarity 26.0%; Pred. No. 4.9e-19;
Matches 143; Conservative 81; Mismatches 228; Indels 99; Gaps 25;
Db 19 LALIFGAVPPAARASAPGSLRAVYHMTTPSGWLCDDP-----ORPVTTHGAYLYLHLS 72
Db 7 LFLLAGAAKISAMTNETSDRPLVHFTPNKGNWMDPNGLWYDEKDAKH-----LYEQYN 62
QY 73 DON---NGPCGWDBASTTDCGVAETHHG--TVMLPLRDPDPVWMSGAVVGTANTAGFGAGAV- 127
Db 63 PNDTVWGCTPLFWGHATSDLTNMEQDPIALPKRNDSSGAFSGSMVVDYNTSGFFNDTID 122
QY 128 -----VALATQPTDGVKRYOEQYLWSTGGFTFTA--PDVIVNTDGRAATTPAEIENA 181
Db 123 PRQRCVAIWTYNTP---ESEEQYISYSLDGGYTFTEYQKNPVL-----ANS 166
QY 182 EWRDRDKIHWDTARGEWVCVIGRLR--YAAFTYSPNLRDWTLRNFDYPNHALGG--IEC 237
Db 167 TQFRDPKVFVWEPYEQKWIIMTAASKSQDYKIEIYSSDDLUKSWKLESF--ANEGLGYOYEC 224
QY 238 PDLFEITA--DDGTRHNVLAASMDAYGIGLPMTYA---YWTGTWDEGQFHADDLTPOWLD 292
Db 225 PGLIEVTEQDPSKSYWMEFISINP---GAPAGSFGNQYFVGSFNGCTHFEAFDQSRVVD 281
QY 293 WGDWYAAVTPSIDAPETKRLAIAMNNKYAARDVPTDASDGYNGQNSIVRELRL--- 349
Db 282 FGKDYALQTFNTDPTYGSALGIAWASNNWEYS-AFVPTNP---WRSSMLSRKFSLNT 337
QY 350 --ARQPGGWTLTSTPVAALTN-----YVTATTLTPDRD--VDGSVLPWNGRAYIEL 399
Db 338 YQNPETELINLKAEPILNISNAGPSWRSFATNTTLTKANSYNVD-----LSN:GTLEFEL 393
QY 400 DIAMDATNNGVSG-----RSPDGTNRHNTIGKYGADLYVDRGPSDLAGYSLAPY 449
Db 394 YVAVNTQITTSKVFADLSLWFKGLEDPPEYLRMGPEVSASSFFELDRGNSKVKFVKNPY 453
QY 450 --SRAAAPIDPGARSVHLRI-----LVDTOSVEVFNAGHTVLSQOVHFAEGD----- 495
Db 454 FTRMNVNNOFPKSENDSLKYKYVGLDQNLILEYFNDDGVSTNTYFMTTGNALGSVM 513
QY 496 -TGIS--LYTD 503
Db 514 TTGVDNLFIYD 524

RESULT 8
US-10-369-493-22749
; Sequence 22749, Application US/10369493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO: 22749
; LENGTH: 581
; TYPE: PR
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-22749
Query Match 12.2%; Score 348; DB 6; Length 581;
Best Local Similarity 24.9%; Pred. No. 4.2e-18;
Matches 132; Conservative 82; Mismatches 226; Indels 90; Gaps 22;
QY 28 PPAARASAPGSL-----RAYHMTTPSGWLCDDPQRTTHGAYLYY 69

Db 53 PFVNTATNGTCLGNYEYLSGYNATDRPKIHFTSPSSGFMNDPNGLVYTGCVYHMF 112
QY 70 LHSQ--NNGPGGWDAHSTGDGVAF-----THHGTVMPLRDFPFVWGSASVVGTA 117
Db 113 QYSPKTLTAGVHGHGTVSKDLIHENYPIAIVPDEHENGVLSL-----PFGSASVVDVH 167
QY 118 NTAGGAG-----AVVALATQPTGVKRYQEQYLYWSTDGGFTFTALPDPVIVNTDGRA 171
Db 168 NSSGLFSNDTPEERIVLIYTDHWTGA--ERQAIATYTDGGYTFK-----KY 213
QY 172 ATTPAEIENAEFRPKIHWDTARGWCVCVIGRLRY--AAFTSPNLRDKTLRRNEDYPN 229
Db 214 SGNPVLIDNSLQFROPKVYDWDANRWNVIVAMSONYGIATFYSSYDLIHWTLSVSTSG 273
QY 230 HALGGTDFEITADDCGRH--WYLAASMDAYGIGLPM---TYAYWTGTWDGEQFHADD 285
Db 274 YLGQYECFPMARVPVEGTDEYKWLFIINP--GAPLGGSVVQYFVGDWNGTNEVPDD 330
QY 286 LTPQWLDGWGDWYAAVTPSIDAPETKRLAIAMNNKYAARDVPTDASDGYNGQNSIVR 345
Db 331 GQTRFVLDKDFYASALYHSSA--NADVIGVGWASNMWQY--TNOAPTQV---FRSAMTVAR 385
QY 346 ELRLARQPGWT---LLSTP--YAALTN--YVTATTLTLPDRTVDSAVLPWNGRAYE- 396
Db 386 KETLRDVPONPMNTLSLIQTLNVSLRDELTLFTAPVINSSSLSGSPITLPSNTAFEE 445
QY 397 -IELDIAMDATNVGISVGR---SPDGR---HTNICKYGADLVDRGSDLAGY-SL 446
Db 446 NVTLISINTEGCTGYCLGRIIISDDPYRLOSISVDVDFASTLVINRAKQMGHNSL 505
QY 447 APYSRAAPIDPGARSVHLRIILVDTQSVEFVFNAGHTVLSQQVHFAEGDT 496
Db 506 FTFSEANDIYIG--NVTLIGIVDGLLELYVNGEKTNTDFFFLQAT 553

RESULT 9
US-10-369-493-10616
; Sequence 10616, Application US/10369493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10616
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Sphingomonas aromaticivorans
US-10-369-493-10616

Query Match 11.0%; Score 313.5; DB 6; Length 311;
Best Local Similarity 30.1%; Pred. No. 7.7e-16;
Matches 101; Conservative 48; Mismatches 135; Indels 51; Gaps 18;

QY 40 RAVYHMTPPSGWLCDPQRPTVTHGAYQLYLYHS--DQNGPGGWDAHSTTDGVAFTTHGT 97
Db 1 RPLXHYAPAAWNLSDPNGLVWDGHEWLFYQYNPFGEDGHSWGHVYSEDLYTWQELPV 60
QY 98 VMLPRDFPVGSGAVGTANTAGAGAVVALAT--QPTDGVKRYQEQYLYWSTDGGFTF 156
Db 61 ALAEEEDGMIFSGSAVIDHQSAGFGKGAWAVYTGARTD--RAHQFQSAASTDGRFT 118
QY 157 TALP-DPVIIVTDGRAATTPAEIENAEFRPKIHWDTARGEW--CVIGRLRYAAFTYS 213

Db 119 TKFTGNPVL-----DLQAD-PRDPNVFVHGPSGRWMSVVLSENRAQLYAS 165
QY 214 PNLRDWTLRRNF---DYPNHALGGIECPDLFEITAD--DGTRHWVLAASM--DAYGIGLP 266
Db 166 VDLRHWDLDSIGRDCGAPGHLW---BCPMWELPVEGTDETR-WLFKVDVLSGAAGQASP 221
QY 267 MYAYW-TGTWDEGEQFHADDLTPQ--WL--DWGWDYAAVTPSIDAPETKRLAIAMNN 321
Db 222 -----WRVGHFGRFP-----VPETGWAVGDHGPDFYAAIGWNAAPDAEGRPVWIGWAGN 271
QY 322 WKYAADVPTDASDGYNGQNSIVRELRLARQPGGW 356
Db 272 HAYQKYLEP-----QGMRGAMTLPRRVSLRDESGY 302

RESULT 10
PCT-US03-07858-1502
; Sequence 1502, Application PC/TUS0307858
; GENERAL INFORMATION:
; APPLICANT: Monsanto
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: PCT/US03/07858
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1502
; LENGTH: 670
; TYPE: PRT
; ORGANISM: Lolium perenne
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (69)..(69)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
PCT-US03-07858-1502

Query Match 10.7%; Score 304; DB 1; Length 670;
Best Local Similarity 24.3%; Pred. No. 1.2e-14;
Matches 148; Conservative 74; Mismatches 243; Indels 144; Gaps 26;

QY 4 AISRRAVLOGAGAGALALIFGGAVPPAARASAPGSLRAVYHMTPTPSGWLCDQRPVYTHG 63
Db 97 AAESGGLGADAGSNAPFWSNMLQWQRTG-----FHFQPEKNMNMNDPMPYVYKG 148
QY 64 AYQLYVLSHDONNGPG---GWDHASTDGVAFTHHGTVMPLRDPD-----VMSGSA 112
Db 149 WTHLFYQYNPEGATWGNKIANGHAVSRDMLRWRHLPIAM-----FPDQWYDINGAWSGA 203
QY 113 VVGTAANTAGGAGAVVALATQPTDGVKRYQEQYLYWSTDGGFTFTALPDPVIVN---TDG 169
Db 204 TV-----LPDGRIVMLYTGSTNA--SVQVQCLAFPSDPS-----DPLLTNMTKYEG 247
QY 170 RAATTPAEIENAEFRDPKLIHW-DTARGEWCVCIG-----RLRYAIFYSPNLRDWTLLR 223
Db 248 NPVLVPPPHVHGKDFRDPPTTAWYDGSWMRIVIGSKDNRRAGMALTYKTKNPHDFELVP 307
QY 224 NFDYPNHALGGIECPDLFEI-----TADDCGRHWVLA 255
Db 308 GVLHVRPATGMCEIDLYPVGGARGIDTAEVAASNSGGGEVHLVHMKESSDDDRH---- 363
QY 256 ASMDAYGIGLPTYAYTWTGTWDEGEQFHADDLTPQWLDGWWDYAAVTPSIDAPETKRLA 315
Db 364 ---DYVALG---RYDAATNKWTPLDADADVGLRLYDWG--KEYASKTF--YDPAKKRRVL 414
QY 316 IAWMNNKYAARDVPTDASDGYNGQNSIVRELRLARQPGGWYLLSTPVAALTNVYATT 375

Db 415 WGVGETDSRADV-----AKGWASLQSPRTVVDLTGTGS--NLIQWPVVVEVETLRTNST 468
QY 376 TLPDRTVDGSAVLFWN-GRAYEIE-----LDIAWDATATNVGIVSGRSPDGTHTNI 425
Db 469 NLGSIIVHSGSVPLSLHRTQDIEASFRDLPLDVAARKEADVGYNCSTSGGAAGRGL 528
QY 426 GKYG-----AD-----LYVDRG-PSDLAGYSLAPYSRAA-----453
Db 529 GPGLLVADARRHGGDTQTAFFVYVARGLDGNLRTFHCHDESRSRANDIVKRVVGN 588
QY 454 APIDPGARSVHLRLTVDTSQVEFVNAGHTVLSQOVHFAE---GDTGISLYTD--GGPAH 508
Db 589 VPVLGD-EALSVRLVDHSIVESFAQGRSVTSRVYPTTEAIYANAGVYLFNNATGARVT 647
QY 509 FTGIVVREI 517
Db 648 ATSLVWHEM 656

RESULT 11

US-10-289-757-176
; Sequence 176, Application US/10289757
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jercoen
; APPLICANT: Forster, Richard L
; APPLICANT: Gibson, John Bryan
; APPLICANT: Shenk, Michael Andrew
; APPLICANT: Norriss, Geoffrey
; APPLICANT: Glenn, Matthew
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use
; FILE REFERENCE: 11000.10610
; CURRENT APPLICATION NUMBER: US/10/289,757
; PRIOR FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: 60/337,703
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 176
; LENGTH: 670
; TYPE: PRT
; ORGANISM: Lolium perenne
US-10-289-757-176

Query Match 10.7%; Score 304; DB 6; Length 670;
Best Local Similarity 24.3%; Pred. No. 1.2e-14;
Matches 148; Conservative 74; Mismatches 243; Indels 144; Gaps 26;

QY 4 AISRRVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDDPQRPVTHG 63
Db 97 AASESGMLGADAGSNAPFWSNMLQWRTG-----FHFQPEKNMNDPNGPVYKG 148
QY 64 AYOLYLHSDONNGPG---GWDHASTTDGVAFTTHGTVMLRPDFF-----VWSGSA 112
Db 149 WYHLFYQYNEGAIGNKTAHGAHSRDLMLRWHLPIAM-----FPDQWYDINGAWSGA 203
QY 113 VGTANTAGFAGAVVALATQPTDGVKRYQOYLWSTDGGFTTALPDPIVYN---TDG 169
Db 204 TV-----LPDGRIVMLYTGSTNA--SVQVQCLAFPSDPS-----DPLLTNWTKEYG 247
QY 170 RAATTPAEIENAEWFRDPKIH--DTARGEWVCVIG-----RLRYAAFTYSPNLRDWTLLR 223
Db 248 NPVLYPPPHVGEKDFRPTTAWYDGGDMWRIVIGSKDNRNRRAGMALTYKTKNFHFDELVP 307
QY 224 NFDYPNHALGGICBDFE------TADDCGTRHWLA 255
Db 308 GVLHVPATGMWECIDILYPVGGARGIDMTEAFAAASNSGGGEVLHVWKESSDDDRH---363
QY 256 ASMDAYGIGLPMYAYWTGWDGEQFHADDLTPQWLDGWDWYAAVTWPSIDAPETKRLA 315
Db 364 ---DYALG---RYDAATNKWTPDLADADVGIGLRYDNG-KFYASKTF---YDPAKRRVL 414

QY 316 IAWNNMKYAAARDVPTDASDGYNGSQNSIVRELRLARQPGGWYLLSTPPVAALTNVYVATT 375
Db 415 WGVGETDSRADV-----AKGWASLQSPRTVVDLTGTGS--NLIQWPVVVEVETLRTNST 468
QY 376 TLPDRTVDGSAVLFWN-GRAYEIE-----LDIAWDATATNVGIVSGRSPDGTHTNI 425
Db 469 NLGSIIVHSGSVPLSLHRTQDIEASFRDLPLDVAARKEADVGYNCSTSGGAAGRGL 528
QY 426 GKYG-----AD-----LYVDRG-PSDLAGYSLAPYSRAA-----453
Db 529 GPGLLVADARRHGGDTQTAFFVYVARGLDGNLRTFHCHDESRSRANDIVKRVVGN 588
QY 454 APIDPGARSVHLRLTVDTSQVEFVNAGHTVLSQOVHFAE---GDTGISLYTD--GGPAH 508
Db 589 VPVLGD-EALSVRLVDHSIVESFAQGRSVTSRVYPTTEAIYANAGVYLFNNATGARVT 647
QY 509 FTGIVVREI 517
Db 648 ATSLVWHEM 656

RESULT 12

US-10-389-566-1502
; Sequence 1502, Application US/10389566
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1502
; LENGTH: 670
; TYPE: PRT
; ORGANISM: Lolium perenne
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (69)..(69)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-389-566-1502

Query Match 10.7%; Score 304; DB 6; Length 670;
Best Local Similarity 24.3%; Pred. No. 1.2e-14;
Matches 148; Conservative 74; Mismatches 243; Indels 144; Gaps 26;

QY 4 AISRRVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDDPQRPVTHG 63
Db 97 AASESGMLGADAGSNAPFWSNMLQWRTG-----FHFQPEKNMNDPNGPVYKG 148
QY 64 AYOLYLHSDONNGPG---GWDHASTTDGVAFTTHGTVMLRPDFF-----VWSGSA 112
Db 149 WYHLFYQYNEGAIGNKTAHGAHSRDLMLRWHLPIAM-----FPDQWYDINGAWSGA 203
QY 113 VGTANTAGFAGAVVALATQPTDGVKRYQOYLWSTDGGFTTALPDPIVYN---TDG 169
Db 204 TV-----LPDGRIVMLYTGSTNA--SVQVQCLAFPSDPS-----DPLLTNWTKEYG 247
QY 170 RAATTPAEIENAEWFRDPKIH--DTARGEWVCVIG-----RLRYAAFTYSPNLRDWTLLR 223
Db 248 NPVLYPPPHVGEKDFRPTTAWYDGGDMWRIVIGSKDNRNRRAGMALTYKTKNFHFDELVP 307
QY 224 NFDYPNHALGGICBDFE------TADDCGTRHWLA 255
Db 308 GVLHVPATGMWECIDILYPVGGARGIDMTEAFAAASNSGGGEVLHVWKESSDDDRH---363

QY	256	ASMDAYCIGLPMTYAYWTGCTWDCGEQHQHADDLITQWLDWGDWDTYAATWPSIDAPETKRIA	315
Db	364	---DYALG---RYDATNKWTPLDADADVIGLRYDWG--KEYASKTF--YDPAKKRRVL	414
QY	316	IAMNNHKYAAROVPTDASDGYNGQNSIVRELRLARQPGGWYLLSTPVAALINYVTATT	375
Db	415	WGWGVEDSERADV---AKGWASLOSIPRTVVLDTKTGS--NLIOWPVVEVETLRTNST	468
QY	376	TLPRVTVDGSAVLPMW--GRAYEIE-----LDIAWDATINVGISVGRSPDGTBHTNI	425
Db	469	NLGSIIIVEHGSVPFLSLHRATQLDIEASFRLLDPLDVAAAKEADGVNCGTSGGAACRGAL	528
QY	426	GKYG-----AD-----LYVDRG--PSDLAGYSLAPYSRAA-----	453
Db	529	GPFGLLVLADARRHGGTEQTAVFYVARGLDGNLRTHFCHDESRSSRANDIVKRVVGIN	588
QY	454	APIDPGARSVHLRLIVDTQSVFVFNAGHTVLSQQVHFAE---GDTFGISLYTD--GGPAH	508
Db	589	VPVLDG--EALSVRVLVDHSIVESFAOGRSVVTSRYVPTETAIYANAGVYLFNNATGARVY	647
QY	509	FTGLIVREI	517
Db	648	ATSLVWHEM	656

```

RESULT 13
US-10-369-493-3411
; Sequence 3411, Application US/10369493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3411
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-369-493-3411

```

[illegible]

```

QY   301 VTWPSIDAPETKRLAIAMWNW---KYAARDVPTDASDGYNGONSVIRELRARQPGGWY 357
      || : || : || | | | | | : | : | : || | | | | : | : |
Db   279 TQNFVDDA--GRNLAVTWIINNNTTKWPSR-----VNGWAGTQSIMREL-FIREDGG-- 327
      || : || : || | | | | | : | : | : || | | | | : | : |
QY   358 TLLSTPVAALTNVTATTTLPDRTPVDCGSAYLWPNGRAYEIELDIAWDTA---TNVGISVG 414
      || || : || : || : || : || : || : || : || : || : || : || : ||
Db   328 -LGOKPV----EEVSLLASGPAKSLGOTDFV---GTYEVGFT---DTARLQATVDLAAS 375
      || : || : || : || : || : || : || : || : || : || : || : || : ||
QY   415 RSPDGTRHTNIGKYGADLYVDRGPS-----DLAGYSI-----APYSRA-----AAPIDPG 459
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   376 DAP-----AFTIELFASGAESTSLVYNFNANRSILITDMNAGYGOAGTWEATIAKPA 426
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY   460 ARSVHLRIILVDTOSVEVFVNAHGHTVLSSQQVHFAEGDTGISLYTDGG-PAHFCTGIIVVREIG 518
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   427 DNRLSLDILDRSIEVFVFGDGTAMTVRVPPRIOESKNIRITSDBGKKTVDFDIVLTPMG 486
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY   519 QA 520
      |
Db   487 SA 488

RESULT 14
US-10-289-757-87
; Sequence 87, Application US/10289757
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen
; APPLICANT: Forster, Richard L
; APPLICANT: Gibson, John Bryan
; APPLICANT: Shenk, Michael Andrew
; APPLICANT: Norriss, Geoffrey
; APPLICANT: Glenn, Matthew
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use
; FILE REFERENCE: 11000.10610
; CURRENT APPLICATION NUMBER: US/10/289,757
; CURRENT FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: 60/337,703
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 87
; LENGTH: 668
; TYPE: PRT
; ORGANISM: Lolium perenne
US-10-289-757-87

```

```

RESULT 14
US-10-289-757-87
; Sequence 87, Application US/10289757
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen
; APPLICANT: Forster, Richard L
; APPLICANT: Gibson, John Bryan
; APPLICANT: Shenk, Michael Andrew
; APPLICANT: Norriss, Geoffrey
; APPLICANT: Glenn, Matthew
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use
; FILE REFERENCE: 11000.1061U
; CURRENT APPLICATION NUMBER: US/10/289,757
; CURRENT FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: 60/337,703
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 87
; LENGTH: 668
; TYPE: PRT
; ORGANISM: Lolium perenne
US-10-289-757-87

```

```

Query Match      10.3%   Score 294;   DB 6;   Length 668;
Best Local Similarity 24.0%;   Pred. No. 7e-14;
Matches 146;   Conservative 76;   Mismatches 243;   Indels 144;   Gaps 26;

QY      4  AISRAVILQAGAGACALALIFGGAVPPAARASAPOSLRVAVHMTPSGWLCDPQRPVTHG 63
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      95  AASESGMLGADAGSNAPFNSNAMLQWRTG-----FHQPEKNMNDPNGPVYVK 146
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      64  AYQLYLHSDQNNGP---GWDHASTTGDGVAFTHHGTVMPLRPDPF-----VWSGSA 112
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      147  WYHLFYQNFEGATWGNKIKANGHAVSRDMLRWRHLPTAM-----FPQWDYDINCANSGSA 201
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      113  VVGNTANTAGFGAGAVVALATQPTDGVRYKYEQYLYWSTDGGFTETALPDPVIVN---TDG 169
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      202  TV-----LPDGRIVMLYTGSTNA---SVQVQCLAFPSDPS-----DPLLTNWKYVEG 245
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      170  RAATTPAIEIENAEWFRDPKTHW-DTARGEWVCVIG-----RLRYAAPTSPNLKDWTLRR 223
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      246  NPVLVPPPHVGEKDFRDEPTTAWYDGDSDGMRRIVIGSKDNRRAGMALTYKTKNFHDFELVP 305
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      224  NFDPVNALHGIECPDLFEI-----TADDGTFRRHWLVA 255
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      306  GVLRHPATGWEECIDYLVPGGARGIDMTAEVAASNSGGGGEVLHVWKESDDDRH----- 361
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      256  ASMDAYGIGLPMTYAYWTGTWGDGQFHADDLTPQWLDWGDWDTAAVWTWPSIDAPETKRLA 315
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

```
Db 362 ---DYAALG---RYDAATNKTPLDADADVGIGRLYDMG-KFYASKTF--YDPAKKRRLV 412
QY 316 IAWNNKYYAARDVPTDASDGYNGQNSIVRELRLAROPGGWYTLTLLSTPVAALNTNYVTATT 375
Db 413 WGVGETDSEADV---AKGWSLQSIPTVVLDTKGS--NLIQWPVVEVETLRTNST 466
QY 376 TLPDRTVDSGSAVLPMN-GRAYEIE-----LDIAMDATNVGISVGRSPDGTTRHNI 425
Db 467 NLASIIVEHGSVPFSLHRRATQLDIEASFRLLDPLDVAAAKEADVGYNCSSTGGAAGRGAL 526
QY 426 GKYG---AD-----LYVDRG-PSDLACYSLAPYSRAA----- 453
Db 527 GPFGLLVLADARRHGGTGTATVFFYVARGLDGNLRFHCHDESRSRANDIVKRVVGN 586
QY 454 APIDPGARSVHLRLVDTQSVFVNAAGHTVLSQOVHFAE---GDTGISLYTD--GGPAH 508
Db 587 VPVLGD-KALSVRLVAHSIVESFAQGRSVVTSRVYPTETAIYANAGVLYFNNTATGARVP 645
QY 509 FTGIVVREI 517
Db 646 ATSLVVKM 654
```

RESULT 15

```
PCT-US03-07858-1977
; Sequence 1977, Application PC/TUS0307858
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: PCT/US03/07858
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1977
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Lolium temulentum
PCT-US03-07858-1977
```

```
Query Match 10.2%; Score 291; DB 1; Length 677;
Best Local Similarity 23.8%; Pred. No. 1.2e-13;
Matches 139; Conservative 66; Mismatches 221; Indels 158; Gaps 25;

QY 40 RAVYHMTPPSGWCLDQRPVTHGAYQLYLHSDQNGPG-----GWDHASTTGGVAF 92
Db 134 RTGFHFQPEKNWMDPNPVYKGNTHLFY----QNPDAIWNKNIANGHAVSRDLVRW 189
QY 93 THRGTYMPLRPDP-----VWGSADVGTANTAGFAGAYVALATQPTDGVRYQE 144
Db 190 RHLPIAM-----FPQWYDINGAMSGSATV-----LPDGRIVMLYTGSTNA--SVQVQ 235
QY 145 YLWSTGDTGFTFALDDPVIVN---TQGRAATTPAEIENAEWFRDPKIHV-DTARGEWVC 200
Db 236 CLAFPSDPS-----DPLLTNNKTYEGNPFVLPYPPPHIGEKDRPDTTAWYDGSQGMRI 288
QY 201 VIGRL--RYAAF---YTSNLRDWTLLRRNFDYPNHALGIECPDLFEITADGTRHWLA 255
Db 289 VIGSKDRHAGMALTYKTNNFIDFELIPGLVHRVPATGMWECIDLIPVGA----- 338
QY 256 ASMDAYGIGLPMTYAYWTGTWGEOPHA-----DDLTPW----- 290
Db 339 ---ARGIDMTEAFAAASNNNGGSEVLHVNMKESDDDDRDHDIYALGRYDAATNKKWTPLDADA 394
QY 291 -----LWGWHDWYAAVTPWPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNCQNSI 343
```

```
Db 395 DVGIGRLYDMG-KFYASKTF--YDPAKKRRLVLMGWGETDSEADV---AKGWSLQST 447
QY 344 VRELRLAROPGGWYTLTLLSTPVAALNTNYVTATTLPDRTVDSGSAVLPMN-GRAYEIE----- 398
Db 448 PRVVLDNKTGS--NLIQWPVVEVETLRTNSTNLGSIPTVVDHGSIFPLSLHRRATQLDIEAS 505
QY 399 -----LDIAMDATNVGISVGRSPDGTTRHNIKYG-----ADLYVDRGPSDLAGYSIAPY 449
Db 506 FRLDPLDVAAKEADVGYNCSSTGGAAGRGALGPFGLLVLDADARRHGGDTE---QTAVY 561
QY 450 SRAAAPIDPGAR-----SVHLRLVDTQSVFV 478
Db 562 FYVARGLDGGLRTHFCHDESRSRANDIVKRVVGNVVPVLDDGEALSIRVVDHSIVESFA 621
QY 479 NAGHTVLSQOVHFAE---GDTGISLYTD--GGPAHFTGIVVREI 517
Db 622 QGGRSVVTSRVYPTETAIYANTGVLYFNNTATGARVTSIATSLAVHEM 665
```

Search completed: May 6, 2003, 12:43:28
Job time : 61.6125 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 02:37:06 ; Search time 5273.5 Seconds
(without alignments)
11042.896 Million cell updates/sec

Title: US-09-868-328B-2
Perfect score: 2001
Sequence: 1 gcggtagcccccagcttcccc.....atcgacagctctctcgctgtt 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1566	78.3	1566	1	AF181254
2	977	48.8	1890	1	AF294730
3	945.2	47.2	1752	1	AB001984
4	931.8	46.6	1551	6	E22826
5	909.6	45.5	1452	6	E22825
6	370.4	18.5	1863	1	AB062062
7	147.6	7.4	65140	6	AX211705
8	147.6	7.4	123580	1	AF263912
9	147.6	7.4	123401	6	AX211739
10	141.8	7.1	3030	8	VCA429230
11	119.8	6.0	3849	6	AX058889
12	119.8	6.0	12744	1	AF204401
13	114.2	5.7	22449	1	SPSNBCDE
14	114.2	5.7	22449	1	SPSNBCGEN
15	112	5.6	77457	1	AF210249
16	111.8	5.6	2951	1	AF123319
17	111.8	5.6	43147	1	SC4A10
18	109.2	5.5	27541	6	AX211706
19	106.8	5.3	30000	6	AX250262
20	106.8	5.3	88063	1	BSZ94043
21	106.8	5.3	209510	1	BSUB0018
22	106.4	5.3	113193	1	AF357202
23	105.2	5.3	15424	1	AE007019
24	105.2	5.3	58930	2	AC098321
25	103.8	5.2	47852	1	MTV023
26	103.6	5.2	14860	1	AE007093
27	103.6	5.2	63033	1	MTV008
28	102.4	5.1	303091	2	AC084799
29	102.2	5.1	298166	2	AC087563
30	101.6	5.1	5994	1	SYSNBDE
31	101.4	5.1	42655	1	SC7H2
32	101	5.0	15311	1	AE007164
33	99.8	5.0	4068	1	ANU12274
34	99.8	5.0	37245	1	SC5F2A
35	99.8	5.0	40790	1	MTCY493
36	99.4	5.0	220469	2	AC074307
37	99	4.9	15348	1	AE007163
38	98.6	4.9	67200	1	MTV017
39	98.6	4.9	77534	1	AF235504
40	98.2	4.9	110000	2	LMFLCHR32_06
41	97.8	4.9	34108	1	SCE25
42	97.6	4.9	150754	9	AC023491
43	97.2	4.9	10420	1	AE005127
44	97	4.8	4615	1	STMWHB12X
45	96.6	4.8	46166	1	SC8F11

ALIGNMENTS

RESULT 1
AF181254
LOCUS
DEFINITION Arthrobacter ureafaciens levan fructotransferase (lfta) gene,
complete cds.
ACCESSION AF181254
VERSION AF181254.1 GI:8163731
KEYWORDS
SOURCE Arthrobacter ureafaciens
ORGANISM Arthrobacter ureafaciens
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Micrococineae; Micrococcaceae; Arthrobacter.
REFERENCE 1 (bases 1 to 1566)
AUTHORS Song,K.B., Rhee,S.K. and Yoo,E.J.
TITLE Nucleotide sequence of levan fructotransferase gene (lfta) from

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Arthrobacter ureafaciens K-2032
Unpublished
21 (bases 1 to 1566)
Song, K.B., Rhee, S.K. and Yoo, E.J.
Direct Submission
Submitted (27-AUG-1999) Microbial Metabolic Engineering RU, Korea
Research Institute of Bioscience and Biotechnology, Taejon 305-600,
South Korea

FEATURES
source
gene
CDS

Location/Qualifiers
1..1566
/organism="Arthrobacter ureafaciens"
/strain="K-2032"
/db_xref="taxon:37931"
/country="South Korea"
1..1566
/gene="lftA"
1..1566
/gene="lftA"
/function="diffructose dianhydride IV (DFA IV) formation
from levan"
/note="levan-assimilating enzyme"
/codon_start=1
/transl_table=11
/product="levan fructotransferase"
/protein_id="AAP73829.1"
/db_xref="GI:8163732"
/translation="MTPAISRRVLQGGAGALALIFGAVPPAARASAPGSLRAVYH
MPPSWLGDPPQVPTTGHAYOLYLHSDNPPGWDHASTDDGVAFPHGTVMPLR
PPFVWSAGVANTTAGFGAGAVVALATOPTDGVRYQEOYLWSTDDGFTFTALP
DPVINTDGRANTPAEINAEWEFRPKIHWDTARGEWVCVIGRLRYAFYISPNLRD
WTLRNFDPNALGGIECPDLFEITADDGTRHWLMSDAYIGLPMPTAYWTGW
DGEQFHADLLTPQWLDWMDWYAAVTPSIDAPETKRLAIAWMNNWKYAAADVPDAS
DGYNGONSVEIRLAROPGWYTLISTPVAALTNVYATTLTPORTVDSGLVLPWG
RAVEIEDLAWDTATNGVISVGRSPDTRHTNIGKYADLYVDRGPSDLAGYSLAPYS
RAAPIDPGARSVHLRLILDVTSQVEVFVFNAGHTVLSQVHFAEGDTGISLYDGGPAH
FTGVIREIGQAI"

BASE COUNT 268 a 635 c 458 g 205 t

Query Match 78.3% Score 1566; DB 1; Length 1566;
Best Local Similarity 100.0%; Pred. No. 1.1e-160; Indels 0; Gaps 0;
Matches 1566; Conservative 0; Mismatches 0;

QY 361 ATGACGCGGCCATCTCACGCGCGCGGTGCTCCAGGAGCGCGCGGAGCACTCGCC 420
DB 1 ATGACGCGGCCATCTCACGCGCGCGGTGCTCCAGGAGCGCGCGGAGCACTCGCC 60
QY 421 CTGATCTTCGGCGGTGCTGCGCGGTGCTGAGCGCGGCGCATCCGCTCGGCGCTCGCTCCGT 480
DB 61 CTGATCTTCGGCGGTGCTGCGCGGTGCTGAGCGCGGCGCATCCGCTCGGCGCTCGCTCCGT 120
QY 481 GCGGTCTACACATGACGCGCGCGGTGCTGCGACACCCCAACGCCCGGTCAACC 540
DB 121 GCGGTCTACACATGACGCGCGCGGTGCTGCGACACCCCAACGCCCGGTCAACC 180
QY 541 ACCACGCGCGCTACCAAGTGTACTACTGCACTCCGACACAGAACACGCGCGCGCGG 600
DB 181 ACCACGCGCGCTACCAAGTGTACTACTGCACTCCGACACAGAACACGCGCGCGCGG 240
QY 601 TGGGACACCGGACGACGACGCGGTGCGGTGCGGCGGTGCTGCGGACCGCGGACCGG 660
DB 241 TGGGACACCGGACGACGACGCGGTGCGGTGCGGCGGTGCTGCGGACCGCGGACCGG 300
QY 661 CTGCGCGCGCGCTACCGCGGTGCGGTGCGGCGGTGCTGCGGACCGCGGACCGGACCG 720
DB 301 CTGCGCGCGCGCTACCGCGGTGCGGTGCGGCGGTGCTGCGGACCGCGGACCGGAC 360
QY 721 GGGTTCGGCGCGCGCGGTGCTGCGGCGGTGCGGACCGCGGACCGCGGACCGGACG 780
DB 361 GGGTTCGGCGCGCGCGGTGCTGCGGCGGTGCGGACCGCGGACCGCGGACCGGACG 420
QY 781 TACCAGGACGATACCTCTACTGTCAGCGCGGCGGTTCAGGTTACCGCGCGCTGCC 840


```

1170 TCCGCGCTTTGCAATATGCACTCCAGCACACCTTCGGAGCGCACAGTCAACGG 1229
QY 1512 CAGCGCGCTTCCTGCAATGAAGCGGCGCATACAGAGATCGAGCTCGACATCGCCTGGGA 1571
Db 1230 CAGTTTCGTACTTCCTGAGGCGGCGGCGGTATGAACATGGAACATCGATATTTTCATGGGA 1289
QY 1572 CACCGGAGAACTCGGCGATCTCGGTGGCGCGCTCCCGCGAGCGGAACCGGCAACAGAA 1631
Db 1290 CACGGCAGGAACGTGGAGATCTCGGTGGCGCGCTCGTCCGATGGCAGCGCATACGAA 1349
QY 1632 CATCGGCAAGTACGAGCAGACCTGTACGTGACCGAGGAGACCTCGGAGCTCGCGGGTA 1691
Db 1350 CATCGGCAAACTAGGTGAGAGAGTTGTACGTGATCGCGATCGCGATCTCGGAGCAAGCGGTTA 1409
QY 1692 CTCGTGCGCCCTTACTCTCGAGAGCGCGCGCCCGCATCGACCGCGCGCGATCGTGCA 1751
Db 1410 TCGGCTGGCACTTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1469
QY 1752 CTTGCGCATCTCGTTCGACACCGAGCGTTCGAGGTTCCTGTAACCGCGCGCGCACACCGT 1811
Db 1470 CTTGCGCATCTTGTAGACACCGCAAGTGTTCGAGGTTCGTAAATTCCTGGGCGCACACGT 1529
QY 1812 GCTCTCCGAGAGTCCACTTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1871
Db 1530 GGTTCGCGAGAGGTGCACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1589
QY 1872 CCGCGCGCGCACACTTCACCGCGCATCTGCTCGCGCGAGATTGGCCAGCGATCTAGCGAT 1931
Db 1590 CGGTGCGGCACTTCACCGCGATCACCATCCGCGAGTTCGCGGAGCGCGCGCGCGCGCG 1649
QY 1932 GCACACACACCGCTCACCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1975
Db 1650 CGTCCACCGCTGGAAGGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1693

RESULT 4
E22826 LOCUS E22826 1551 bp DNA linear PAT 18-JUN-2001
DEFINITION Levan fructotransferase gene.
ACCESSION E22826
VERSION E22826.1 GI:13024125
KEYWORDS JP 1999069978-A/2.
SOURCE Arthrobacter nicotinovorans.
ORGANISM Arthrobacter nicotinovorans
Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales;
Micrococciaceae; Micrococciaceae; Arthrobacter.
REFERENCE 1 (bases 1 to 1551)
AUTHORS Masakazu, S., Atsushi, Y. and Fusao, T.
TITLE Levan fructotransferase gene
JOURNAL Patent: JP 1999069978-A 2 16-MAR-1999;
NIPPON OIL CO LTD
COMMENT OS Arthrobacter nicotinovorans
PN JP 1999069978-A/2
PD 16-MAR-1999
PF 28-AUG-1999 JP 1997232421
PR MASAKAZU SAITO, ATSUSHI YOKOTA, FUSAO TONITA
PC C12N15/09, C07K14/195, C12N1/21, C12N9/10, A61K38/45, (C12N15/09,
PC C12R1:06),
PC (C12N1/21, C12R1:19), (C12N9/10, C12R1:19), C12N15/00, A61K37/52,
PC (C12N15/00, C12R1:06)
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source i. .1551
FT Location/Qualifiers
FT 1. .1551
FT /organism="Arthrobacter nicotinovorans".
FT /db_xref="taxon:29320"
BASE COUNT 315 a 524 c 443 g 269 t
ORIGIN

```

```

Query Match 46.6%; Score 931.8; DB 6; Length 1551;
Best Local Similarity 75.8%; Pred. No. 3.6e-92;
Matches 1184; Conservative 0; Mismatches 367; Indels 12; Gaps 2;

QY 361 ATGACGCGCGCATCTACAGCGCGCGCTGCTCCAGGAGCGCGCGCGAGACACTCGCG 420
Db 1 ATGAGCTATGACATCTCTCGCGCACTGCGCTGCAAGGTGCAAGGTTGTTGCTTGGCA 60
QY 421 CTGATCTTCGCGGCTGCTGCGCGCTGCGAGCGCGGCGCATCGCTCCGGGCTCGCTCGT 480
Db 61 CTTTTCATGAGCAATGCGCATTCCTGCGCGCGCGCA-----CGCCAGGCAATCCCTCCGG 114
QY 481 GCGCTCTACACATGACGCGCGCGCGCTGCTGCGAGCGCGCGCGCGCGCGCGCGCGCG 540
Db 115 GCGATCTACACATGACGCGCGCGCGCTGCTGCGAGCTGCTGATGTCGCGAGCGCGTACAT 174
QY 541 ACCAGGCGCGCTACAGCTGTACTACTGTGCTGCGAGCGCGCGCGCGCGCGCGCGCGCG 600
Db 175 ACAAGCGCGCGCTACAGCTGTACTACTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGGA 234
QY 601 TGGGACCGCGGAGCAGCAGCGCGCGCTGCGCTTCACGCGCGCGCGCGCGCGCGCGCG 660
Db 235 TGGGACCGCGGAGCAGCAGCGCGCGCGCTGCTTACACCGCGCGCGCGCGCGCGCGCG 294
QY 661 CTGCGCGCGACTTCCCGCTGTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
Db 295 ATGCAACCGGACTTCCCGCTGTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 354
QY 721 GGTTCGCGCGCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Db 355 GGTTCGCGCGCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 408
QY 781 TACGAGGAGCTACTCTTACTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
Db 409 TTCAGGAACAGTACTCTTACTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 468
QY 841 GACCCCGCTCATCTCAACACCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 900
Db 469 GACCCCGCTCATTGTGACACTGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 528
QY 901 GCGAGTGTTCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
Db 529 GCAGATGTTCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 588
QY 961 GTATCGGAGCTGCGGTACCGCGCGGTCTACCTGCGCGCGCGCGCGCGCGCGCGCGCG 1020
Db 589 GTATCGGAGCGCGCGCGCTGCTGCTTCTACACCTCTCCCAACCTCGCGGATTTGGCAA 648
QY 1021 CTTGCGCGCAACTTCGACTTACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
Db 649 TGGAGTCCCACTTCGACTTACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 708
QY 1081 TTCGAGATCACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
Db 709 TTCGAAATGACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 768
QY 1141 TACGCGCTCGCGCTCCCGATGAGCTACGCTGAGCAGCGCGCGCGCGCGCGCGCGCGCG 1200
Db 769 TACGCGCTCGCGCTCCCGATGAGCTTGTGCTTACTGAGCAGGTTTCATGGAACCGCGCA 828
QY 1201 TTTCCAGCGCGAGCACTCACCGCGCAATGGCTGCGACTGGGCTGGGCTGGTACGCGCGCG 1260
Db 829 TTCATCGCGGAGCACTCACCGCGCAATGGCTGCGACTGGGCTGGGCTGGTACGCGCGCG 888
QY 1261 GTCACCTGCGCGATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320
Db 889 GTGACCTGCGCGCGCGTGGAGCAGCTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 948
QY 1321 AACTGGAAGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Db 949 AACTGGAATATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAA 1008

```



```
QY 1426 TGGTACACCTCTGTAGACACCCCGTGGAGCGCTGACGAAGTACGTACCGCCACACACC 1485
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 TGGTACACCTGTGTACGACGCGCGTTCGGCGCTTTCGAATATATCCACCTCCAGCACC 1014
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1486 ACACCTCCCGACGGACCGTTCAGCGCAGCGCGTCTCCATGAGCGGCGCATAC 1545
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1015 ACCCTTCGGAGCGGACAGTCAACGCGTTCGTACTTCCTGGAGCGCGCGGTAT 1074
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1546 GAGATCGAGCTCGACATCCCTCGGACACCGGACGAGCTCGGATCTCGGTGGCGCG 1605
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1075 GAACCTGGAAGTATGATATTCATGAGACGCGGACGAGCACTCGGTGGCGCG 1134
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1606 TCCCCCGAGAACCGCGGACGACGACATCGGCAAGTACGAGCAGACATGATCGTAC 1665
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1135 TCGTCGAGTGGAGCGCGCATACGAACTCGGCAATACGAGTGTACGTCGAT 1194
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1666 CGAGGACCTCGACCTCGCGCGGTACTCGTCCGCGCTACTCGGAGCGCGCGCGCC 1725
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1195 CGCGCATCTCGGAGCAAGCGGTATGCGCTGGCACCTACACCGCGCGCGCGCC 1254
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1726 ATCGACCCCGCGCGGATCGCTGCGACCTGCGATCTCGTCGACACCGAGCGTACG 1785
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1255 ATCGATGCGAAGCGGATCGCTCGACCTGCGCATCTTGTAGACACCCAAAGTGTGAG 1314
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1786 GTCTTGTCAACCGCGGACACCGCTGCTCTCCAGAGGTCCACTTCGCCGAGGCGGAC 1845
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1315 GTGTTCTGTAATTCGGGCGACAGGTGTTCGACGAGGTGACGAGTGTACGTCGAT 1374
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1846 ACGGGAATCTCGCTTACACGACGCGCGCGCGCGGACACTTACCGGATCTGTCGTCG 1905
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1375 ACGGGATCTCGCTTATCGGACGCGGCGGTTCGGGCAACTTACCGGATACCATCCG 1434
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1906 GAGATTGGCGCGCGATC 1923
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1435 GAGTTGGGAACCCCATC 1452
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 6
AB062062
LOCUS Microbacterium laevaniformans mRNA for levnanase, complete cds.
DEFINITION Microbacterium laevaniformans (strain:ATCC 15953) cDNA to mRNA,
ACCESSION AB062062
VERSION AB062062.1 GI:14289196
KEYWORDS Microbacterium laevaniformans
SOURCE clone.lib.pLEV-KB.
ORGANISM Microbacterium laevaniformans
REFERENCE 1
AUTHORS Song,E.K., Kim,H., Sung,H.K. and Cha,J.
TITLE Cloning and characterization of a levnanohydrolase from
JOURNAL Microbacterium laevaniformans ATCC 15953
MEDLINE Gene 291 (1-2), 45-55 (2002)
PUBMED 12095678
REFERENCE 2 (bases 1 to 1863)
AUTHORS Song,E. and Cha,J.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-2001) Eunyoung Song, College of Natural Science,
Pusan National University, Department of Microbiology; San 30
Jangjun-Dong Gumjung-Gu, Pusan 609735, Korea
COMMENT (E-mail:eksong95@hanmail.net, Tel:82-51-510-3328)
Submitted through BRIC(Biological Research Information Center) of
korea
URL: http://bric.postech.ac.kr
GeneNuri No. KS106615.
FEATURES
1. .1863
Location/Qualifiers
/organism="Microbacterium laevaniformans"
/strain="ATCC 15953"
/db_xref="taxon:36807"
```

```
/clone_lib="pLEV-KB"
/notes="re_sites:kpn1, SacI;vector:pBluescript;average
insert size:7931 bp-vector lab_host:E.coli"
1. .1863
CDS
/codon_start=1
/transl_table=11
/product="levnanase"
/protein_id="BAB59060.1"
/db_xref="GI:14289197"
/translation="MGRPPADGRRRRPGGCARTLTTTEGSGDDAPALPARRALLA
AGLALAGSAAALVAGPAAEDDAPSLRHFVTPDHMNDPQPVVLDVHLVYLL
NYADTDANFRANFTGWRLEATSDGVAFADQAPKGTNANYDLWSGAVVDHAGTA
GFGPVMVLTQMDHPHAAQKLDASQQAQFLWYSVDGRTFRDGPVVDGRR
LRDPKVVDDEERQVALIAERDRVSEYTPDLHRMTGTGTECPDLFR
FDADGTTWVGWSANGATNEPATYATWTSFGDSFVDPVAAPWLDHGFQWYGA
VTWDPAPLELRRIALAWNNWDYAHGAPTRDGTGTDSITRIARAGAGVSL
SAPVDALQDHATHHTALGDVRVDGTELAYRBDAYEITRISRODADNVGQLRSAD
GSRHADGLTRDIYALNRAQTGRPSWRVESRPLDDADTVELRLVDRRTIEVFVG
GRTHSSQVFAPSGDGLALYTSGGPAVFRDLRITFEALAHVPRASRTSRATGAR
GRRPPAPSGSVRRRAICAGRRARAWPTRSPGTRPRPASRRRPSRSTGTRCTS"
BASE COUNT 270 a 728 c 627 g 238 t
ORIGIN
Query Match 18.5%; Score 370.4; DB 1; Length 1863;
Best Local Similarity 55.7%; Pred. No. 1.5e-31;
Matches 948; Conservative 0; Mismatches 691; Indels 63; Gaps 10;
QY 301 CGCTGCCCGAACACGGGTGACGCTTCGTGCGCGCCACCCGTCACGAGGAGAACACGCA 360
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34 CGCGCGCGCGGAGGTGCGCGCGCCGCTGACACACACAGAGAGAGATCCGAT 93
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 361 ATGACGCGCGGCGCAT---CTCAGCGCGCGCGCTCTCAGGAGCGCGCGCGGACGACATC 417
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 GACGCTCCGCGCGCTGCGCGCGCGCGCTGCGCGCGCGCTGCTCCCGCTCGCGCTC 153
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 418 GCGCTGATCTTCGCGCGGTCTGTGCGCGCTGACGCGCGCGGATCTCGCTCCGGGCTCGCTC 477
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 GCGCGCTCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 213
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 478 CGTGGCGCTACCATGACGCGCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCTC 537
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 CGCGCGCGGTATCATCTTCAGGTGCGCGGACCACTGGATGAAGACACCGCGCGCGCGCTC 273
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 538 ACCACCGCGCGCGCTTACCACTGTACTACTCTACACGCGCGCGCGCGCGCGCGCGCG 590
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 274 GTGATCGATGCGCTGCTGCACTACTACTCTACACGCGCGCGCGCGCGCGCGCGCG 333
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 591 -----CCCGCGCGGTGGACACGCGGAGGACGCGCGCGCGCGCGCGCGCGCGCGCG 639
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 CGTGGCACTTCGGCACCGAGTGGCTCTCGCGGAGCGCGCGCGCGCGCGCGCGCGCG 393
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 640 CACCGCGCGCGCGGTGATGCGCGTGGCG-----GCCCGCGCTTCGCGTGGTCCGGTGG 693
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 394 GATCAGGGGTCGCGCGCACCCAGGCGACGACGACCACTAGACCTCTGCTGCGGGTGG 453
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 694 GCGGTCTCGCGCACCGCGCAACACGCGGTTTCGCGCGCGCGCGCGCGCGCGCGCGCG 753
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 454 GCGGTCTCGACCATGCGCGGAGCGCGCGGTTTCGCGCGCGCGCGCGCGCGCGCGCG 513
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 754 ACCCGCGCGCGCGCGGTGCGCGAAGTACAGGACGAGTACTCTACTGCTGCGCGCGAC 813
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 ACCCGAGATGGACCA-----CCCCACCGCGCGCGAGAGCTCGACGCGTTCGCGCGCG 567
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 814 GCGCGGTTCACGCTTCACCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 568 GCGCGGTTCCTCTGGTACTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 627
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 874 GCGCGCGCGCGCGCGGATCGAGACGCGCGGATGGTTCGCGCGCGCGCGCGCGCGCG 933
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 628 CCGGTTCATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 681
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 934 GACCGCGCGCGCGGAGATGGGTCTGCGCTATCGGACGAGTCTCGCGTACCGCGCGCTTAC 993
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 682 GACGACGAGCGCCGCTGGGTGCGCCCTCATCGCCGAGCGGAGCCGGTTTCGTCTAC 741
QY 994 ACCTCGCCGAACTTGGCGACTGAGCACTTCGCGCAACTTCGACTACCCGAAACACGACC 1053
Db 742 ACCTCGCCGAACTTGCACCGCTGAC-----CGGACGGCGAGTACGTCAACGCGGC 795
QY 1054 CTGCGGGCATGAGTGGCCCGGACCTTTCGAGATCACGCGAGACGAGGACACGCCAC 1113
Db 796 ATCGGACGATGAGTGGCCCGGACCTTTCGCGCTTCGGGCGGACGACGCGACGCCAC 855
QY 1114 TGGGTGCTCCCGCCGACGATGAGCGCTAGCGCATCGGCTTCCCTGAGCTAGCGCTAC 1173
Db 856 TGGGTGATGGGCTCAGCGGAAAGGATACGCGACGAACCGCGACCTACGCGTAC 915
QY 1174 TGGACAGGCACTTGGGACGCGGAGTTCACGCGCGGACGACCTCACCGCGCAATGGCTC 1233
Db 916 TGGACGGCTCGTTCGACGGGTGACCTTCGTCGCCGAGCTGCTGCGCGAGTGGCTC 975
QY 1234 GACTGGGGCTGAGTGGTACGGCGGCTGACCTTGCATGAGTGGCGGCGGCGGAGACC 1293
Db 976 GACACAGGGTTCGACTGGTACGGCGGGTACCTGGGAGGACCGCGCGGACGCTCGAG 1035
QY 1294 AAGCGCTCGCCATCGGTGATGAACACTGGAAGTACCGCGGACGCTCGCCACC 1353
Db 1036 CGCGGCTCGATCGGCTGATGAACACTGGGACTA---CGGACGGGCGGCGACC 1092
QY 1354 GAGCGATTCGACGGCTACAAAGGCGAGAACTGATCGTCGCGAGCTGCGGCTCGCCGA 1413
Db 1093 TGGCGGGGACGAGTTCACCGGACCGACTCGATCAGCGGCGAGATCCGCTCGCCG 1151
QY 1414 CAGCTGGGGCTGATACACCTTCGTAGACACCGCGTGGCGAGGCTGAGCAACTAGTC 1473
Db 1152 --TGGCGTTCGGGCTACAGCTTCTCGCTTCCGCTCGCGTTCACCGAGTTCGCTACCG 1209
QY 1474 ACCGCGACACCACTCCCGGACGCGCTGCGAGCGGAGCGGCGCTGCTGCCATGAAC 1533
Db 1210 ACCGACAGAGCGGCTCGGCGAGCTCGCGTTCGAGGGTTCACCGAGTTCGCTACCG 1269
QY 1534 GAGCGGCTATACGATTCGAGTTCGATCGCTGGGACACCGCGAGCGGCTCGGCATC 1593
Db 1270 GGGACGCTACGATACGACGCGCATCTCGCGCAGGACCGGACAGCTCGGGTTC 1329
QY 1594 TCGGTGGCGCTTCGCGGAGGAAACCGGACACGAACTCGGCAAGTACGAGGACGAC 1653
Db 1330 CAGCTCGGCGCTTCGCGGAGGACGCGGACGCGGCGGCTTCACCGCGGCTAC 1389
QY 1654 CTGAGCTGACGAGGACCTTCGAGCTTCGCGGCTGCTGCTGCGGCTTACTCGGA 1713
Db 1390 GCCTACTCAACCGGGGACAGCGGGGCGGCGGCTGCTGGAAGG-----TCGAGA 1441
QY 1714 GCGCGCGCGGATGACCGCGGCGGCTGATCGGCTGATCGGCTGCTGCTGCTGCTGCTG 1773
Db 1442 GCGCGACGCGCTTGA---CGAGCGGACACGCTGGAGCTCGGCTGCTGCTGCGGCG 1497
QY 1774 CAGAGGCTGAGGCTTCGTCAGCGCGGCGGACACCGTGTCTCCAGAGGTTCACTTC 1833
Db 1498 ACAGCATGAGGCTTCGTCGCGGAGCGGCGGCTACAGCACTTCAGGAGGTTTCGCG 1557
QY 1834 GCGGAGGCGGACAGGGAATCTCGCTTACACGAGCGGCGGCGGCGGCTTCACCGG 1893
Db 1558 CCGTCGCGGCAAGGCTTCGCGCTTACACCTCGGCGGCGGCTCGGCTTCGCGGAT 1617
QY 1894 ATCGTCTGCGGAGATGGCGAGGCTTATAGGCGATGACACCAACGCGCTTCACCGG 1953
Db 1618 CTCGGATCAGCGAGTTCGAGACCTTCGCGAGCTTCGCGGCGGCTTCCTCGGAGTTCG 1677
QY 1954 GCGCGCGCGGAGAGGAGG 1975
Db 1678 AGGCGGACAGTGGGCGGCGG 1699

LOCUS AX211705 65140 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 1 from Patent WO0159126.
ACCESSION AX211705
VERSION AX211705.1 GI:15523937
KEYWORDS Streptomyces noursei.
SOURCE Streptomyces noursei.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 65140)
AUTHORS Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
Gulliksen,O.M.
TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 1 16-AUG-2001;
Norwegian Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
(NO)
FEATURES
source location/Qualifiers
1..65140
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"
BASE COUNT 8270 a 25171 c 22273 g 9426 t
ORIGIN
Query Match 7.4%; Score 147.6; DB 6; Length 65140;
Best Local Similarity 45.4%; Pred. No. 5.7e-08;
Matches 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;
QY 8 ACCCGGACTTCCCTCGAGACACCGTCCCTACCGCGGACCGCGCGCGGCGGACTGC 67
Db 19803 ACGGCGGAGACACCGCGCGCTTCGGACTGACCGCGGCGGCTTCGAGCGCGCAC 19862
QY 68 TCTCAGCTAGACGGCGGCGCTCTCGAGGCTTCGTCGGGGACGCTGAGCGACTGCGT 127
Db 19863 AACACCGCGCGCTACGCGGACCTCGGCGCATACGCGCGGCGGCTTCGCGCT 19922
QY 128 CGAACCTGCTCTTCGCGGCGGCGGCTGACCGGAGGCTTCGAGACGCGACGCGCAG 187
Db 19923 GGAAGCGCTCTCGTTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 19982
QY 188 GAACCTGTCAGCTGACGCGGCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 247
Db 19983 CCGCGGAGACACCGCTACCATCGCGCTACGCGCGGCGGCGGCGGCGGCGGCG 20042
QY 248 CTGCGCGGCTTCGCGGCTGACGAGCGCTCCGCGGCGGCGGCGGCGGCGGCGGCG 307
Db 20043 TCGACTCTCTGCTCTCCGCGAGGTCGCGCGGCGGCGGCGGCGGCGGCGGCGG 20102
QY 308 CGAACCGGCTGAGCGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 367
Db 20103 ACCGCGACTCTCTTTCACGCTGAGTGGACCGCGGCGGCGGCGGCGGCGGCGG 20162
QY 368 CGGCGCATCTCAGCGCGGCGGCTGCTCCAGGAGCGGCGGCGGCGGCGGCGGCGG 427
Db 20163 CGGCGA-----CCGTCGCGGCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGG 20213
QY 428 TGGCGGCTGCTGTCGCGGCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 487
Db 20214 TCGCGG-----CCAGCGGATCGGAGCACCGCGGCGGCGGCGGCGGCGGCGG 20264
QY 488 ACACATGAGCGGCGGCGGCTGCTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 547
Db 20265 CGGACCGGAAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20321
QY 548 GCGCGCTACGAGCTGACTACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 607
Db 20322 GCGCGCGGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 20381


```

/codon_start=1
/transl_table=11
/product="Nysg"
/protein_id="AAF71763.1"
/db_xref="GI:8050837"
/translation="MASPDLEERTAPARRLLGCLLRPHRRSVALAVSMGVGIVL
NAFGLLLGRVDLLDAGVGLGVPAGCIDFAAGRLLLVLLAYVVALSFLAGGR
LVASAVRTIHELRRDAREKRLRLPHRFDRQAPAGELLKRTINDINLOOTLAE
LITSFSLTMLVLMGVISPLAVMLLSVPVSAITAARISKRAOPHYAAQSMANGTL
NARVECTCHALIKGIDRRRAAEERFDANDAVYRAAKAQFASGAMPVMMFVANL
GVYLVAGIAMKVINLTGLDGVQAFILYARQFISOPITELIASVAGRLQSGIASORVF
TLLDAPEQADPLRCPCTPARAEGRVEFTDVSFRYSPDPTLJENLSITVPEGSTVAIVG
PTGACKTTLGNLLMRPEYDPDSGRILLDGTDTATWTRDDLSRFLGLVQTLWTFPGTTIA
ENIYAGACACRADEIEAARATCAORFIRTLPGQYDITLDDESGTVSAGEKOLLTVAR
AFLARPAVLVDLSESVSDTRTEVLFIQRANSLRAGRTSFIARHLSTIRDAALLIVM
DAGRIVEOCTHQDLCAOGLYARLHAARTHTPTAGAAAG"
complement(2662..4416)
/gene="nysg"
CDS
complement(2662..4416)
/gene="nysg"
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="NysH"
/protein_id="AAF71764.1"
/db_xref="GI:8050838"
/translation="MLLRLLRAQLRPYAWATAALVALQVLQILGTLTLLPTLGAALIDQ
GVYRGDGRITELGVVMGVVVALVOIAAALGAALAAATAMGRDRLRSALFRILDFS
AREIGRFGTSLTTSVNDVQVONLAQTGFIYVCAPLMCLGSLVLLALRQDYPALLL
LVALNVAVCGLLARMGTLIYARMLTDLRLGLRLREAITGVRRVRSFVRDHERA
RFAQNDAPLVYVRVGRGLIATPVLVLLMNGFTVALLWTSGRHIDAGRMPIGSLSA
ELYSLLTMSVYLAFLVFLSVPRARVCAGRIAEVLDGTSSVAPPAQPPVRGPAGRI
LSCAAGRYGAEPEEVLVDLTVEPGERIAVLGSGTKTLLNLVRLADATGAV
RVGTDVRELTAAALAAVGFVPPRPLFSGVTVASNLRRGRDPDATDEELWEALRAQA"
ADFAVMPDGLDAETLQGGNVYSGGQRQLSARALLRRPEIYLFDDCFESALDQNTDA
ALRTALPYTAGATVITVAQRISAGRDADRIIVLDGRVVAAGTHTDVLVLTSTPTREI
ALSQLTEEAAGHLAGR"
4714..5748
/gene="nysDIII"
4714..5748
/gene="nysDIII"
/function="presumably involved in mycosamine biosynthesis"
/note="putative dGDP-mannose-4,6-dehydratase"
/codon_start=1
/transl_table=11
/product="NysDIII"
/protein_id="AAF71765.1"
/db_xref="GI:8050839"
/translation="MSKRALITGITQDGSYLAEHLLSQGYQVWGLIRGOANPRKSRV
SRLASELDFIDGDLMDQGLSVSAVDTPQDEVYNLGAISFVPMWQOAELEVTVNGMG
VLRMLEAIRMVSGLSTSRVSPRGQIRFYQASSEMFGKAAETPQRETTLPHRSPYG
AAKAYGHVITRNYRESFGMYAVSGLMFNHNESPRRGQEFVTRKISLAVARIKGLQDLK
ALGNLDVARDMCGYAGDYVRAMHLMQODAGDDYVIGTGOMHSVRDARVRAPEHVGILNW
EDYVVIDPLVRPAEVEVLCADSAQDLRLCKPKVDVPTLNRMMVDSDLAQVSRNO
YGDVLLAANW"
5930..34363
/gene="nysi"
5930..34363
/gene="nysi"
/function="responsible for condensation steps 9 to 14 in
the nystatin polyketide backbone synthesis"
/note="polyketide synthase"
/codon_start=1
/transl_table=11
/product="Nysi"
/protein_id="AAF71766.1"
/db_xref="GI:8050840"
/translation="MDNEQKLRDYLLKLATADLRFRPRRVHKLESAQEPVAILIGMTCR
YPGQSPEDLWRMYEAGEHGVTPEPTDRGMDLEALAAAPTASGGLFHDAPDFADFF
GISPREAVAMPQQRVLESAAWEAFERAGIDPYSLSRTGVIFIGAMAQDYRVGPDG
AEGFLTNTGTSLSGRISYTFGTVGPAVTVDTCSSSLVAVHLATQALRAGECTLAL
AGGVTIMSGPGTFIEMGRQGLSADRCRCSFGDADGTGMAEGVGLVLERLSDAVRN
```

```

GHEILAVRGCTAVNODGASNGLTAPNGPSQOQVIOQALVHARLAAGDIDVVEAHGTGT
TLGDDPVEAALLATYQGNRPADRPDLILGSKNLSHTQAAAGVAGVVKWMAMRRHTL
PRTLHAEPTTHVDWSQAVRLLTDTDPWPCAPARRAAVSFGISGTAHTLIEQAP
EPQPEDATQAODDAAAGSTPATPVVGVVLLSGRTPDALRGDQAAAALRAALDRRP
DILDLAHSATTRAGFEHRAVLATDHPALTQGLTALADDDPAAAPAWITGTTRAET
RLAVLFTQGOAQRILGAGRELAARFAPAFATLDAALDAFTPHLDRLPREVLWGTDALL
RATYQAQPALFAVEVALYRLIESFVRPHLAGHSVEIIVAAHLAGVLSLDAATLVA
AKRMLQALPDGGAMIQAQSEADPAVLGHAGDEQVIAAIVNGSPSLVSGAEATVTA
LAEQAADGCKTRRLRVSHAFSPIMEPLDAFRAVVEDTLQPLLPVYSNLTKGPA
TVAQILTADYWDVHVRHAFADGIDWLARHDTTAPLELPGDGVLSAMAQDCLDAADA
WPKALQSGTADLRSVCLGAHHPILLSAVSLADAGCTLLTGLSRLTQPHLADTVYRG
TTLPGCTAFLELAVRAGDEVGCDRVEELTAAPLLLPEQGGVQVQLMIGNPNVSGRRT
VNWHRPDTGDDTPTAHATGVLTTADASRQLPASSEGGTFLAGDPHDAADAAQPP
AGAELPLDGHYDLRLADGGFGYGVFGLRAAMRGDGVYIAEVELPEAGRSDAEAFGL
HPALLDAALHAAFPFTGLGERGGGLPFSGWESVLSHAGGATLRLVRLTPVADALATV
ADGTGAPVLSVDSLVRISVATQOOLDTAAVARDALFRIDWTPVPTATDPGVVALIGA
DPFGLLTHAGFADAPAYPDLAALAADGPVTIVVLSLACTGDDAADPARSAHRCAR
ALAAVOTLWIDHHERFAARLVFTRGATVGRDVAANAAMVGLVRSQSENPFGCFALVDL
DPDGAAGAAALVAALVSGEPQLAVRGDVLRLVARLRRPLTEVAGADGDDGVDGSG
VSFSEGAADVLTGTGGGLGAVLARHLVAEYGVDRDLLVSRGERAVGAGELVAELAGV
GARVRVACDVTDRAAVVELGVGHASVAVHAAGVLDLDDGMGALTGERLSAVLRPKVD
AVMHLHEATRGDLDADFVFSVLGAGVSGPGOANYAANAFLDALMTRRAREGLPGLS
LANGPWEQSGGNTGLTLDVDAERLARSGVPLSVAQGLALFDAAVAGTDTATCPVRLD
LPVLRAGEVPTGLLSLIRVARRAAVAGSAGTANLAQRLRDEDEGDEMVDLUVRG
QVALVILHATGGDVGDAFRADFDLSLTAVELNRNLNTVTLRLPATLIFYDPTVRH
LATYVLDELGTDAEVATQPAAVADDPIVIVGMACRYPGVSGSPDLRWLTETGT
DVGFPNTRMGDVESLHPDHPHGTSTYRSGLFHEAGFDPGFGFMSPREALATD
SOORLLLESSWEAIEFRAGIDPVLSRGTGVFAGVMSYDYSAMLASPEFEGQSGSS
PSLASGRVAYTLGLSGPVTVDATCSSSLVAMHMAQALRSCEGLLAGGVYVMSPT
AVFDFQRGLSPDGRCAADAAGVGWSEGVVLVERSDAVNRNGHELIAVVRG
SAVNDGASNGLTAPNGPSQOQVIRQAASGLGTAGDVIDVEAHGTGTGLDPIEQAQ
LIATYCRDREPERLLLSVKNLGTQAAGVAGVVKMVLAMRHGVPRTLHDAAPS
SHVDNSAEPELLSEQAAMPETGVRVRAGVSFSGISGTVHVIVEQAPCAKIAAAGA
ARTPCAVPVLSSGRSALRGOAARLHGLQAPDAELVDVALSIATTRSREORAA
VYRTQDLIASIGLALADRPVAVVEGEAAGRTAVTLFTCGQSORAAARHELHEVO
PEFAAFDAVCAVDFPLDRLPREVVEEDGSDERALLDETGTQTPALFAVEALFRL
VESWGRDVFVAGSIGETAAAHVAGVLTLEDAECLRLAARATLMQALPTGGAMIAIQA
TEDEIAHLDDTVAIAAVNGPSQVVISGDEAAETIATFAERGTKRRLRVSHAFHS
PRMDGLDAFRIVAEGLTVRAPRIPLVSLDTRRADDAEVCYAEYVVRHREAVRPAD
CVRTLDRAGATTFLESGDGLTMAEDTLGDHDAELVPLMRAEABEALAAATLAR
LOVRGDVDMWAAVLAGTCARTDLTPTAFQOHAYYWPOLPTPAALAAADPAQOOLWAA
VERGDARELADILGIGEOQLTPLDSILLPALTWSRNGQEKHLLDLTRYVRETRLSKP
TAPVLDGTLVVASDADQADPALLDLADLAGSHGARVRRLLDSCADRAVLAERL
ARTADVDAATQVLSVPLDERDADDCPLTLGLALTVALQALDADTGAQRGLWTATRG
AVSTNPADPVTHPVQAAAGLGRGVLEHPRLGLVLDLTVQFDERAGORLAGILAVK
DAPGCEOVALRTGVSGRRLVRHTVEALPTAAETATGTVLITGCTGLGAEARWL
APAGAQLHLVTSRRGPDAPAAELRAELEGGYSPSVVACDVADDAALAAVLTALPEE
LPLTGTVHTAGVGHYGLDTLSTAEFAGLTAAKLAGAAHLADLADRELDFVLFSGSI
AGVWGSGQSAITGAANAYLDLALHRRRGLAATSVANGPWAEGMAADDAVSETLRR
QGLGLLDPAAMTELRRVAVRODVTTVADVMQRYAPLTSARFSAALTAGLPEVRAL
ADDERTEQDATGSEVTVRVALABPEQLRLLDLTVTESATVLGHSSADAVEGRAE
RDVGFDSLTAVELRKLGAATGLSPVMVFYPTPLEAQYLRAELILGAEVAGPV
ATGGADDEPIATIGMACRPGCVSSPEOLMDIVASGTDASIEFPVNRGWTGHLDPD
PDRPGTITSTQGGFLHEDEPDTFEGISPREALYMDPOORLLLETTHWESFERAGIRP
EFLRSTLTGTFFVGSSTYGTGAGDGTGEGHVMVTSQSVLSRLLSTFVLEGPAVTVD
TACSSSLVALHILACOSLRNGESNLAVAGGATIMTTPNPFIAFSRORALAKDGRCKAFS
DDAGMTLAEGVYVLRSDAQRNHPVLAVLRGSAINODGASNGLTAPNGPSQOQ
VITROALANARLAPGIDIDALEHGTGTPLGDPTEAQLFATYGRDRDPSALLGSVKS
NIGHTQSAAGTASVTKMYMALRHSLEPPTLHADAPSSHVDWAGTVRLLTOARAMPET
GRPRAAVSSGIGISCTNAHVLEQAPVADTPAERPAPVAPVPIAGVVPVWYVTSAA
ALURGQERLLHAETGTALPAAGPLDGLSLVSARARFEHRAVVPVVPAGTDPLEAAL
AVATDGPSPVAVGADVDEGRTVFFPGQSQWGMGSQLDESFAEFAEIAEACAAAL
AEFTDWSLVDLGRVYGAPSLERVVDQFASVAVMSLAALWRSRGLVPDAVYVHSQ
ETAAAVVSGALSRLDGRVVALRQAIGRALAGRGGMMSVALSDVLEPRLVFEFGRV
SVAAVNGPSVSVVAGEPEALDHALHARTADDIRARRIADYASHSHVEDLHEELLEV
LAELNPTSEVPEFSTVTGDMADGYFNRILGRVREPADADLAAEYRAFY
EYSSHPVLSMAVQEAIDACGVPVAAAGTLRDQCGCTDRFLLSAAEFVYRGVDVWAGL
FEGTGAIDRLPTIYAFQHEHLWAPPPAAVAAADDDAAFWTAVEDGDVSALTATAG
TDEDVSAVALPTYSRRRARDPSTVDAMRYAVANKPLGGTLPHSLTGTWLLVATDG
PDDTSVAGALETYGAERVRLVIDEEDVRLRAGAEADVGTIVSLVAAERTDAD
PCTSLVLTALTVALIOALGDAEIDAPVVALTRGAVSTGRADDELTAPOQAVTGTGT
AALEHPQRWGGLTDLPAALDARAQRLAAVLSGALGSDOLAI RFSGVFTRRIVAREA
```

Query Match 7.4%; Score 147.6; DB 1; Length 123580;
Best Local Similarity 45.4%; Pred. No. 4.6e-08;
Matches 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;

Qy 8 ACCCGGACTTCCCTCGAGGACCCGCTCCGCTACCGGCGGACCGCGCGGCGGCGGCTGCG 67
Db 79656 ACGGCGGAGACACGACCCCGCGCTTCGAGCTGCACCCCGCGCTGCTGAGCGCGCAC 79715
Qy 68 TCCTCAGCTAGACGGGCGCTCTCGAGGCTTCCTCGGGAGCGTGAAGCGACTGGT 127
Db 79716 AACAGCGCGCGCTACGCGACCTCGGCGGCATCAGCGGCGGCGGCTGCGGTTCCGCT 79775
Qy 128 CGAACCCTGGTCTCTCGTGGGCGCGGCGGTGTACCGGAGCTCTGAGACGCGACCGCGAG 187
Db 79776 GGAAGAGGCTCTGCTCGCGCGCGCGCGCACACACGCTCCGCGCGCGGATCGCGCCCG 79835
Qy 188 GAACCGTCACTGACCGGATCGAGCTCGAGGCGCCGAGGATGCTACGCGCCCTGAAC 247
Db 79836 CCGGCGAGGACCGCTACCATCGCGCTTACGAGCGCGCGGCGGCGGACCTGTGTCCG 79895
Qy 248 CTGCGCGCTGCTGGGCTGAGAGCGCTCCACCGCGGACAGCTCTCTTACCGCTGCC 307
Db 79896 TCGACTCCTGCTCTCCGCGAGGTCTCCGCGGAGCGACCGCGCGCGCGGACCGCTCC 79955
Qy 308 CGAACCAGGGTGGAGCTTCTGTCGCGCGCCACCGGCTCCAGAGAGAACAGCAATGACGC 367
Db 79956 ACCCGGACTCTCTTCCAGCTCGAGTGACCGCGCTCCAGGCGCGCGCGCGCGCGCAC 80015
Qy 368 CGGCCATCTACCGCGCGCGCTGCTCAGGAGCGGCGCGGCGGAGCAGCTCGCGCTGATCT 427
Db 80016 CGGCA-----CGTCCGCGCTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCACACC 80086
Qy 428 TCGCGGCTGCTGTGCGCGCTGACCGCGCGGATCGCTCGCGGCTCGCTGCGGCTCT 487
Db 80067 TCGCG-----CGACCGGATCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCTCG 80117
Qy 488 ACCACATGACCGCGCGCGCTGCTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 547
Db 80118 CCGAGCGCGGAGGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80174
Qy 548 CGGCTACACGCTGCTACCTGCTGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 607
Db 80175 CGCGCGCGCTCG 80234
Qy 608 ACGGAGCACGACGCGCGCTGCTTACGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 667
Db 80235 AGTGCTCGCGGACGCGCTTGGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80294
Qy 668 CCGACTTCCCGGTGCTGCGGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 727
Db 80295 CCGAGCGGACG 80354
Qy 728 CGCGCGCGCGCTGCTGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 787
Db 80355 AGAACCGCGCGGTTTCCGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80414
Qy 788 AGCAGTACTCTACTGCTGAGCGCGCGGCTTACGCTTACCGCGCGCGCGCGCGCGCGCG 847
Db 80415 AGACCTTGCG 80471
Qy 848 TCATGCTCAACCGGAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 907
Db 80472 GCACCGAGTGCAGCG 80531
Qy 908 GGTTCGCGGACCCAGATCAGTGGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 967
Db 80532 GGAACCGGAGCGGACCGCTCTGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80591
Qy 968 GACGACTCGGTACGCGCGCTTACACTCTCGCGGAACTCGCGGAGCTGGGAGCTTCCGCG 1027
Db 80592 CCGCGCACTGTCGCCACCG 80651

Qy 1028 GGAATTCGATACCCGAAACACCGCCCTCGCGCGCATGAGTGCCCGGACCTGTTTCGAGA 1087
Db 80652 CGGCGCGGAGCGCGCGCGGACGACCTGAGCGGAACTACCGGGCTCGCGCGCAACGTC 80711
Qy 1088 TCACCGAGACGACGCGGACCGCACTGGTCTCGCGCGCGCATGAGCGCGCTACGCA 1147
Db 80712 ACATCGCGCTGCGACGTGCGCGGACCGCGCGCGCTCGCGGACCTGTCGCGACCGTCC 80771
Qy 1148 TCGGCTCTCCCATGACGTACGCTACTGAGAGGACACCTGGAGCGCGGAGGAGTTCACG 1207
Db 80772 CGGCGGCGACCGCTACCGCTCGTCTGCTCCACACCGCGCGCTGCTCGAGCGCGTCC 80831
Qy 1208 CCGAGACCTACCCCGCAATGGCTCGACTTGGGCTGGGACTGTGAGCGGCGCGTCACT 1267
Db 80832 TCGGCTCTCCCTACCGCGCGCGCTGAGACCGCTCTCGCGCGCGGCGCGCGCGCGCT 80891
Qy 1268 GGCCATCATGACGCGCGCGGAGAGCGCGCTCGCCATCGCTGGATGAGCACTGGA 1327
Db 80892 GGCAGCTGCAGAGGCG 80951
Qy 1328 AGTAGCG 1387
Db 80952 TCG 81011
Qy 1388 TCGTCCGCGAGTGGCGCTGCG--CCGAGAGCTTGGGCTGGTACACCTCTCTGAGCAC 1445
Db 81012 ACGGCTCG 81071
Qy 1446 CCCGTGCGAGGCTGAGCACTAGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1505
Db 81072 CGTGACCG 81129
Qy 1506 CGACG 1565
Db 81130 CGCGCGCTCG 81189
Qy 1566 CTGGAGACCGCGGAGCG 1622
Db 81190 CTTGGCG 81249
Qy 1623 GCACAGCAACATCGGCAAGTACGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1682
Db 81250 CGAGGCG 81309
Qy 1683 CGCGCGGCTACTGCTCG 1736
Db 81310 CGCGCGGCTCTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81369
Qy 1737 CGCGCGGCTGCG 1796
Db 81370 CGCGCGCGGGAAGCG 81429
Qy 1797 CG 1856
Db 81430 CG 81489
Qy 1857 GCTCTACACCG 1904
Db 81490 CACCG 81537

RESULT 9
AX211739
LOCUS AX211739 125401 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 35 from Patent WO0159126.
ACCESSION AX211739
VERSION AX211739.1 GI:15523950
KEYWORDS Streptomyces noursei.
SOURCE Streptomyces noursei.
ORGANISM Streptomyces noursei.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 125401)

AUTHORS Zotchev, S.B., Sekurova, O.N., Fjaervik, E., Brautaset, T.,
Stroem, A.R., Valla, S., Ellingsen, T.E., Sletta, H.V. and
Gulliksen, O.M.
TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 35 16-AUG-2001;
Norges Teknisk Naturvitenskapelige Universitet (NO); STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO);
ALPHARMA AS (NO); Sinvent AS (NO); Zotchev, Sergey Borisovich
(NO); Sekurova, Olga Nikolayivna (NO); Fjaervik, Espen (NO);
Brautaset, Trygve (NO); Strom, Arne Reidar (NO); Valla, Svein
(NO)
FEATURES
source Location/Qualifiers
1..125401
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"
BASE COUNT 15664 a 49692 c 42871 g 17174 t
ORIGIN
Query Match 7.4%; Score 147.6; DB 6; Length 125401;
Best Local Similarity 45.4%; Pred. No. 4.6e-08;
Matches 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;
QY 8 ACCCCGACTTCCCTCGAGACACCGTCCCTACCGGCGGACCGCCGCGCCGACTGC 67
DB 80064 AGCGCGGACACCGACCGCGCGCTTCGGACTGACACCGCGCCCTGCTGGACCGCGCAC 80123
QY 68 TCCTCAGCTAGACGGCGCCCTCTCAGGTCTTCGTCGGGACGCTGAGGGACTGCGT 127
DB 80124 AACACGCGCGCTACGCGGACCTCGGCGCCATACGCGCGCGCCCTGCGCTTCCGCT 80183
QY 128 CGAACCTGTCTCTTGGGGCGCGGTGTGACCGGAGCCTCGAGCGGACGCGCCAG 187
DB 80184 GGAAGCGCTCTCGCTCGCGCGCGCGCCACACCGTCCGCGCGCGATCGCCCGG 80243
QY 188 GAACCTGTACGTGACCGCGATGAGTGCAGGCGCCGAGCATGTGAGCGCCCTGAC 247
DB 80244 CGCGGAGGACACCGTACCATCGCGCTGACGACCGCGCGCGACCGCTGCTGTCG 80303
QY 248 CTGCGCGCTTCTGGGCTGACGAGCGCTCCACCGCGAGCTCTCTTCTACCGTGC 307
DB 80304 TCGACTCTCTGCTCCCGGAGTCTCCCGCGAGCGACCGCGCGCGCGCGACCGTCC 80363
QY 308 CGAACAGGGTGGACGCTTCTGCGGCCACCGCTCCACGAGAGAACGACGAATGACGC 367
DB 80364 ACCGCGACTCCCTCTTCCACGCTGAGTGACCGCGCTCCAGGCGCGCGCGCGCAC 80423
QY 368 CGGCCATCTCAGCGCGCGCTGCTCCAGGAGCGCGCGCGCGGACTCGCGCTGATCT 427
DB 80424 CGGCCA-----CCGTCGCGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCACCC 80474
QY 428 TCGGCGGTGCTGTGCGCGCTGACGCGCGCGGATCGCGCTCGGCTCGCTGCGCTCT 487
DB 80475 TCGCG-----CCACCGGCATCGGACACCGCGCGCGCGCGCGCGCGCGCGCTCG 80525
QY 488 ACCATATGACGCGCGCGCGCTGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 547
DB 80526 CGGACCGGAGGCG 80582
QY 548 GCGCTTACAGTGTACTACTGCTGCTGCGACGACGACGCGCGCGCGCGCGCGCGCG 607
DB 80583 GCGCGCGCTGCG 80642
QY 608 ACGGAGCAGCAGCGCGCTGCTTTCACGACACGCGCGCGCGCGCGCGCGCGCGCGCG 667
DB 80643 ACTGGCTCGCGGACGACCGCTTTCGCGGAGCGCGCGCTGTGCTGTGCTACCGCGCGCGCA 80702
QY 668 CCGACTTCCCGGTGTGCGGTGCGGCGTGTGCGGACCGCGCGCGCGCGCGCGCGCGCG 727
DB 80703 CGGAGGACACCG 80762
QY 728 GCGCGCGCGGTCTGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 787

DB 80763 AGAACCCGCGCGCTTTCGCGCTCTCTGACCTCGCCCGCGACACCGCGCGCGCGCGCG 80822
QY 788 AGCAGTACCTTACTGTGTGACGACGCGGGGTTCAGTTTACCGCTTCCCGCGCGCGCGCG 847
DB 80823 AGACCTGGCCACCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80879
QY 848 TCATCTCAACACGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGT 907
DB 80880 GCACGACGTGACGCGCGCGCGCGCTGCGCGCTGCTCCGCTCGCGCGCGCGCGCGCG 80939
QY 908 GGTTCG 967
DB 80940 GGAACCGGACGCGCGCTGCTGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 80999
QY 968 GAGGACTGCGGTACGCGCGGTTCACACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1027
DB 81000 CCGCGCGCGGTGCG 81059
QY 1028 GCAACTTGAATACCG 1087
DB 81060 CG 81119
QY 1088 TCACCG 1147
DB 81120 ACATCG 81179
QY 1148 TCGGCGCTCCCGATGAGTACGCGCTTGTGACAGGCGCGCGCGCGCGCGCGCGCGCG 1207
DB 81180 CG 81239
QY 1208 CGGACGACTCAGCG 1267
DB 81240 TCGGCTTCCCTCAGCG 81299
QY 1268 GGCATCTGATGACG 1327
DB 81300 GGCACCTGACGAGGCG 81359
QY 1328 AGTACCG 1387
DB 81360 TCG 81419
QY 1388 TGTTCG 1445
DB 81420 ACGCGCTCG 81479
QY 1446 CCGCGTGGCG 1505
DB 81480 CGTGGACCG 81537
QY 1506 CGACG 1565
DB 81538 CGCGCGCTCG 81597
QY 1566 CTGGGACACCG 1622
DB 81598 CCGTGGCG 81657
QY 1623 GCACAGGAATCGGCAAGTACGAGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1682
DB 81658 CAGGCGGACATCG 81717
QY 1683 CGCGCGGTACTCGCTCG 1736
DB 81718 CGCGCGGCTCGGACG 81777
QY 1737 CGCGCGATCGGCG 1796
DB 81778 CGCGCGCGCGGAGCG 81837
QY 1797 CG 1856

GLPDEDGURAAAAALLRRHPLNLRAGFWQOQVERPVQVFPNEVPLPWHTRDFALGESD
RERAVEAYTADRAERFDPGAPPLIFALLGLAADHKLVLVTHLLDLDPGSMPLLR
BLFTLYGORGDDGAPVPVTPYRAYLAHLAGDDAARAALVTHLLDLDPGSMPLLR
GRGAGSALPQOITWHEIDAATTAGLTALABSCNITLQHPGAKRLGTAAARPOLGRDGV
VFGATVAHPPEIGESTIGMETINTLPVRVVRPAETLGLLGRVQREQAALIEHRH
LSLDIRSTTGELFDVTVFVFNTPIDPAVLRAEGRRLAGFEVSQATHYPLSLLA
IPFESNRLRHDGVDLDEAGARULLERLTLTDIAEHGADLPVGRLLDLASAERHQ
VLEGNLGTGPRGYDLNAPLAEQAARTPDTTALLVGRSLTYAELNARANLARHLV
TLGAGPEQIVAKIKERSLDLYVALLVLTGCAAYLPVDTYPAERIVAMPMDAPAVLV
LTGDETDGLSGYDQDLDLADRTAPLPAHPAVVITYTSGSTGTPKAVVMPDCAAVNL
LAWHRRETPGAGAPVAGQTTIGFVAGAEILATWLHGKTLAVPSQEVRRSAEQLAAM
LDEQHVSELYAPNLVIEKAAEAAEAGRTLPALRHIAQAGEALTLTRVREFAAAPVG
ROLHNHGVPAETHNMTGALDEEPAANSEHAPLRGPVSGARVYVLDLSALRPAVPGVT
ELYLAGAGVSRGYNRPVLTAEFVADPYAPSGARMTGDLGRWNTGRGELEFAGRA
DHQVIRGFRPEGEIEAALTDLPAVAAAIVVREDRPGDKRLVAYAPAGEIGDAAA
VRSRLATLPDMPVPAIVALDALPLTPNGKLDRAALPAPOSTTAVORTPSREEIL
TTLPAAVLKPRVGDIFDLGSHSLATRLISKIRTVLGAETLRLDFEAPVAGL
GERLDQASGAPALTPAERPERLPSHAKORLWFLGRDGNPNYINIPALRSLTGL
TAALRAALADVAHETLTRYPSHDEGEYQHI VAPADATPALTVHTEHRLAEOLA
EATARPDLTTELPRLATLFRIGDEHVLHLLHHIAGDGWSLAPLTRDLAAAYARR
NRLARHLVAGLGPQEQIVALQPLRSLLDVTAVLAWKSGAAYLPVDPDYPPEERTAMR
ADARPAVWDAIPDOTTAAVADSLRTADRSAPLLPAHPAYVITYTSGSTGAPGVV
AHRSLAATVPAQAAAGLGTSHSRVNFASISFDAAMWELTSALFTGSLADADLL
PGPSRLVHRHITLALPESALPDALPGDTPGTDLIVAGDATAPQOARFAPGR
MVNAYGTTETATMSEPGDGPPIGRPVHAHVTVLDERLRPVPPVGTGEMAYD
GAGVARYLHPALTAQFDPDPYALLFGTETRMVGTDLARLDADGRHLFAGRADQ
QVTRGFRIEGETETLTHPAVAGAVIAREDTPGDKQIVLVTTRTHHAAPDVL
LTKGVDRWOOTYOSLYAEAPGRADEFSCWNSYTAETIPLMEERWRAATVDRL
ALRPRVLEIGSGTLLISOVAHPTEEVYRGTLSRSVVARLAHLGPPDLADKVTVD
ARAHEHTDLPKGRFDITVNSVAQYPPDARYLAGILHRAELLAPGDTIFELDIRNL
RTLHAFRTQNGEALPDLARALPVGVGVAKFDLHLSWVELRGTGDRPGQIAALE
YATDLFRTTAQLDRLRLAAVADDPGLTVGRVELLDPARHRVLDWNTTTPGGD
DITLRTAHNHEMSHRYDVLHRRGSDSPAETVLRWGTDLDTAGELARVLSGPRV
RVTLGPNARLTGETRAQAALERSAADACARLEAPAPAGALDPHDVYALGAVCTWGH
DUTDPAVAGDTA VPRGTROGPLANDPARAEODAAVLGRDLRVAORLPAHPAAY
VLLDRPLSANGKLDRALPADRGEDAAQOAPRTREIILSTLFAELVGLPANGIDR
SFFDLGSHSLATRLSIRSVLGAELAVRDLFOAPVVAALAEVDAQOARPLTAR
TRAEVPLSFAQYRLWELHMEGPGTYNIPMSLRLGTDLTALRAALADVTRARHEA
LTVTPYSSGVPGHVLHVRPFETADTTAOTLEADLTRAARGFDLAREPLRATL
FRLAEDHEVDLHMHIAAGDWSWPLARDLGDAYAAARLAGRAPDLPLPVQYADAL
WORLDGDDDDPTSRHAROLAYWAGALDPEELQPCDRPRPAVATHRGTVPFADV
PDLHQRNALATEGRASLFWLOAFALLTRHAGTDTIPLGSPAGTDDALELVC
FVNTVLVTRDTSQDPTRELLQVRRESDLAAAHQDVPEFKLVEKLRPERSLRHPL
FQVMLAFONSGEARLALPDLARALPVGVGVAKFDLHLSWVELRGTGDRPGQIAALE
YATDLFRTTAQLDRLRLAAVADDPGLTVGRVELLDPARHRVLDWNTTTPGGD
GAARTPVTELRQAARTPOVALVHDGRLTYAELHARANRLARHLITLGCAGPEQIV
ALRMPRSLLLYVALLVLTGAAAYLPVDISYPABRIAFMIEDARPTVTLDRPDLGA
YRDTLTDARTAPLRPEHPAYVITHSGSTGPKAVVMPHAGLANLTHWARRPGET
GVRTAQTAIGFDSVQELISPLVWGTILAVPSEVRHSIELLAGWLETQOINELFAP
NLVLEALAEAAEAGRTLPDITDILQGEALAPTRERAFVTAAPVGRRLHNHYGPAET
HAYTHHTLADAPHAPSPAPIGRPVDRVYVLDLSALRPPVGTGTYLHAGVARG
YLPRETTATREFAVDYAPRGRMYRTGDLGRWNTGRGELEFAGRADHVRKIRGFVE
PGEVEAALTAHRTIOAAVLHAGDLVAVVTTADRLTGVREHLARLPLRPFVPTAY
VRLPALPTVNGKLDRALPADAPAGRAPRTVREELICGLFAEVLTRAGIDDNF
FDLGHSLATRLSIRLRTALGAEITLRLDEAPTVAGLGERLDAQSARPLTAER
PERPLSQAQRLLSRLGTGDPGPNPVLGRVLSGTDLTAAALTAALADLVARHETLR
LFRLEGEHVSQHVLDLPGARALPHRDTTQDDLATELATAAAARHCFDLTTELPRLAT
LFRLEGEHVSQHVLDLPGARALPHRDTTQDDLATELATAAAARHCFDLTTELPRLAT
LWQHALLGENDNDPSTISROSAHMKDALAGPEEIQLPADRPATPAHREAVFTVL
DADHRELLHILRTGASLPMVLOAAVAVLTRHAGTDTIPLGSPAGTDDADLVL
GFVYNTVLVTRDTSQDPTRELLQVRRESDLAAAHQDVPEFKLVEKLRPERSLRHPL
LFQVLLALONTPEAVLEPLGLTARPEVVTGLGAARFDTLNFNAERHDPGDEPHDIDL
EYSTDLEDRNTPEAVLEPLGLTARPEVVTGLGAARFDTLNFNAERHDPGDEPHDIDL
RGRTADACILPRITAEQAARTPAHVAVTEAGTGLITYSELDAANRLARHLIRGVTA
ETRVALLAEERSAQLVTTTALLKAGVYVPLHTGYPVDMRHRVLADTEAALLDTHH
AATAALDITPALVDEDTTAGEODTTPADTVRPOQLAYIMFTSGSTGPKIGITHR
DAIALAADRCDWLDGSRVLMHSYAFEDISTFELFAPLLAGRIVAPRGDIDRAVLQ
RTLAAGHVTSLLLTAGLLGVTADEAPEVFTGVKDVWTVTGGDVSPTAVRRVLEACPGTV

VKTLVGPTEPTTIGCTWLFTDPRIPPAVPIGRPLDNTRAYVLDRLERLVPVPGVTGEL
YIAGAKVGRWDSQARTAEERTADPHALFGDTGGRMYRTGDLARADGVLHFCGR
ADQOVLIRGERTEPEGEIETALAAHDPDTRAAVAPVRAGDKVLVAYLVTPADAGDTT
AEOLRAGLETQLPDYMPAAAFVALPALPVTYNGKLDRLDALPEPWGGGAGRPGRPRE
ELICGLFREVLAGPRVGDIDNFFELGSHSLATRLVGRKTVLADIGVITLFEAPTV
AALAAIDGADTGDPFGVWLPRLTGGSRAPLFCVHPAGGFWGYTSALLRHTDREOPL
YALQARGILARPEPLDDIDAMARDAEQIRKTYPEGYEILGWSFGGLVAHAAARLQ
AEGARVSLLLVLDGYDAYDGTHEVGEQVLAILLNAGVDRAGFGDAPLQRAVL
EKLRRSGSALGNLDDADVMTVFLNLRILIONFRPRFTGDMVFFAAAGHHDHDPAL
TPGNWRPVYTRIEEHLDDTHAGLAEALCAVARTLAQRDTPQAGH"

BASE COUNT 3012 a 10220 c 6589 g 2628 t
ORIGIN

Query Match 5.7%; Score 114.2; DB 1; Length 22449;
Best Local Similarity 43.6%; Pred. No. 0.00032;
Matches 840; Conservative 0; Mismatches 1050; Indels 36; Gaps 6;

QY 33 CGTCCCGCTACCGCGGAGCGCGCGCGGCTGCTTCTCAGCTAGAGCGGCGCGCTCT 92
Db 831 CGACGCGCTGGCGCGCGCGGAGAGCGCGCGGCTTCCCGCGCGCGCTCGC 890
QY 93 CGAGGTCTTCTGTCGGGAGCGGTGAGCGGCTGCTGGAACCTGGTCTCTCTGGGGCGCG 152
Db 891 CGCTACGTCCAGTGTGCTGACGGGAGCGGACCGTCACTCCCTCGGCTGCGGTCAACGC 950
QY 153 CGTGTGTGACCGCGGAGCTCTGAGACGCGGCGGCGGCTGCTGAGCTGACCGGATCGA 212
Db 951 CGCGCGCACTTCCGCGCGCGCGCGCGGCGCGGCTGCTGCGGCTGCGGTCAACGC 1010
QY 213 CGTTCAGGCGCGCGGAGGATGTGACGCGCGCTGAACCTGCGCGGCTTCTGGCTGAGGAGC 272
Db 1011 CGTCAACCGCGCGCGCGCGCGGCTGCTGCGGCTGCGGTGCGGCGCGCT 1070
QY 273 GCTCCACCGCGAGCTCTCTCTTCTACCGTGGCGGCGGCGGCTGAGGCTGCGCGC 332
Db 1071 CGAGCGCTGTCTGGCGGCGGCGGCTTACCGCTGACGAGGCGGCGGCGGCTGCGCGC 1130
QY 333 GCGCCCGCTTCCAGGAGGAAACAGCAATGACGCGCGGCTTCTACGCGCGCGCGTG-- 390
Db 1131 GGTGCGGAAACGGGCTACGGCGCGCTGCTGCAAGCTGATGCGCTACGGCGCGCGCTGGA 1190
QY 391 ---CTTCAGGAGCGCGCGCGGAGGAGTCTGCGCGCTGATCTTGGCGGCTGCTGTGCGCGCT 447
Db 1191 CTTCGCGCGCGCGCGCGGCTGCTGCGCGCTGCTGCGCGCGCGCGCTGAGGAGCTGAA 1250
QY 448 GCAGCGCGGCGATCGCTCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 507
Db 1251 GTTCTCGGTACGAGGAGCGCGGCTGCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1310
QY 508 GGTGCTGCT-----CTGCGCGCGCGCGCGCGGCTTACCGCGCGCGCGCTTACCGCT 560
Db 1311 GCGCGGTACGCGGAGGAGCGGAGCGCGCGCTTCTGCGCGCTTCTGCTGCTTCTGCTGCTG 1370
QY 561 GTACTACCTGCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 620
Db 1371 CGCTTCTGCGGAGGCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1430
QY 621 CGACGCGCTGCTGCTTACGCGCGCGCGCGCGGCTGCGCGCGCGCGCTTCTGCGCGCT 680
Db 1431 CG 1490
QY 681 GTGGTTCGCGGCTGCGGCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 740
Db 1491 CG 1550
QY 741 CGTTCGCGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 800
Db 1551 CGCGGAGCTTCGAGCG 1610
QY 801 CTGGTTCGAGCGGCGGCTTACGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 860
Db 1611 CCCGAGCGGCTGCTGCGCGCGCTTCCCTCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1670

```
QY 861 CGAGGTCGCGCGCCGACACCGCCGCGGAGATCGAGACCGCGGAGTGGTTCCGCGACC- 919
Db 1671 CGCGGCTCTCAAAATCGCGCGCCGCTACCTGCGCGCTCGACCCGCGCTACCGCGCGACCG 1730
QY 920 ---TCCAAGTCCACTGGGACACCGCCCGGAGAAATGGTCTGCTCATCGACGACT 974
Db 1731 GCTGGCTGGATGCTCGAGGACCGGCGCCCGCTTCGTCCTCGACCCGAGCGGATGAC 1790
QY 975 GCGGTAGCGCGCTTCTACACCTCGCGGAACCTCGCGACTGACACACTTCGCGCGCAACTT 1034
Db 1791 CGAGGACCTCGCGCTACCGGACACCGCCCGCGTACCGCGTGCAGCGCCCGCCACCC 1850
QY 1035 CGACTACCGAACACCGCCCTCGCGGCGATCGAGTGCAGCGCCGACTGTCAGATACCGC 1094
Db 1851 CGCGTACGTCATACACCTCGCGCTCCACCGCGCCCGCCCAAGGGGCTGCTGATCCGCA 1910
QY 1095 AGACGCGGACACGACCTGCGTCTCGCGCGCCACGATGAGCGCTACGCGATCGCCT 1154
Db 1911 CACACAGTGGTGGCTGTTCACCTCCACCGGACCACTGGTTGCGTTCGCGCCCGCCGCA 1970
QY 1155 CCCCATGAC---GTAGCGCTTACTGGACGACCTCGGACGGGAGCAGTTCACACCGCA 1211
Db 1971 CGGTGACCGCTCTCCACTCTACGCTTCGACTTCGCTCGGAGATCTGGGGCGC 2030
QY 1212 CGACTACCGCCGCAATGGCTCG-----ACTGGGGCTGGGACTGGTACCGCGC 1259
Db 2031 CCGTCTGACCGCGCGCGCTCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2090
QY 1260 CGTCACTGCGCATCGATCGACCGCGCGGAGACCAAGCGCTCGCCATCGCTGGATGA 1319
Db 2091 CTTCTCGACCTGCTGGCGCGCGAGAGGTCACCGCTCTCAACACGACACCGACCGCCTT 2150
QY 1320 CAATGGAAGTACCGCGCGACGCTGCGCCACCGACGATCGACGGGTACACACGGCA 1379
Db 2151 CCACGAGCTCGACCGCGCGACCGCGCGCTACCGCGCGCGCGCGACTGCGCTCAGTA 2210
QY 1380 GAATCGATGTCGCGGAGCTGCGGTCTCGCGCGCGACAGCTGCGGGTGGTGTACACCTCT 1439
Db 2211 CGTGTGTTGCGGGGCGGAGCTCGACGCTGCGCGCGCTGCTGCGCGGCTGCTGCGCGCG 2270
QY 1440 GACACCGCGTGGCAGCGCTGACGACTACGTCACCGCGCACACACACACACTCCCGAC 1499
Db 2271 CGGACGCGCGCGCGCTGCTCAACATGATACGCGATCACCGGACACACCGCTGACGCTC 2330
QY 1500 GACCGTGCAGCGCGCGCTGCTGCGCATGGAACGCGCGCATGAGATCGAGTCTCGA 1559
Db 2331 GCACGCGCGCTCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCTGCTGCTGCGCGCGC 2390
QY 1560 CATCGCTGGGACACCGCGCGAGCTGCGGATCTCGTGGGCGCG---TCCCGCGCGCG 1616
Db 2391 CATCCCGGACCTCACCGCGCGCGCTGCTGCGCGCGCGCGCTGCGCGCGCGCGCT 2450
QY 1617 AACCGCGCACACGATCGGCAAGTACGAGCAGACCTGTAGTGCACCGAGGACCGCTC 1676
Db 2451 CACCGGGAGCTGATGCTCGCGGGGCGCGCTGCGCGCGCGCGCTGCTGCGCGCGCGCG 2510
QY 1677 CGACCTCGCGGGTACTCGCTCGCGCGCTTACTCGGAGCGCGCGCGCGCGCGCGCGCG 1736
Db 2511 GCTGACCGCGCGCGGTTCCCGCGCGCGCGCTACGCGCGCGCGCGCGCGCGCGCTATCG 2570
QY 1737 CGCGCGATCGGTGACCTGCGGATCTCTGCGACACCGCGAGCGGTGAGGTCTTCTGCA 1796
Db 2571 CACCGCGACCTGTGTCGCGCACCTCGACGACGCGAGTACGCTTACCTGCGCGCGCGCG 2630
QY 1797 CGCGCGCGCACCGTCTCTCCAGAGGTCCACTTCGCGCGGAGGCGGACACGGAATCTC 1856
Db 2631 CGACGAGTCAAGATCCGCGGCTTCCGGATCGACTCGCGGAGATCGAGACGCTCTCGC 2690
QY 1857 GCTCTACACGCGCGCGCGCGCGACACTTCACCGGCTGCTGCTGCGCGGAGATTGGCA 1916
Db 2691 CACCGACCGCGCGCTGCGCGCGCGCGCGCGCTGCTGCTGCGCGGAGGACCGCGCGCGCT 2750
QY 1917 GCGCAT 1922
```

Db 2751 GCGGCT 2756

RESULT 14
SPSNBGEN
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
AUTHORS
TITLE
JOURNAL
REMARK
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

SPSNBGEN 22449 bp DNA linear BCT 24-MAR-1997
S.pristinaespiralis snbc and snbde genes.
X98690.1 GI:1906377
Pristinamycin I synthase 2; Pristinamycin I synthase 3 and 4; snbc
gene; snbde gene.
Streptomyces pristinaespiralis.
Streptomyces pristinaespiralis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 22449)
de Crecy-Lagard, V., Blanc, V., Gil, P., Naudin, L., Lorenzon, S.,
Famechon, A., Bamas-Jacques, N., Crouzet, J. and Thibaut, D.
Pristinamycin I biosynthesis in Streptomyces pristinaespiralis:
molecular characterization of the first two structural peptide
synthetase genes
J. Bacteriol. 179 (3), 705-713 (1997)
97158664
9006024
2 (bases 1 to 22449)
de Crecy-Lagard, V.
Direct Submission
Submitted (21-JUN-1996) V. de Crecy-Lagard, Institut Pasteur, 25
Rue de Dr. Roux, Paris Cedex 75724, FRANCE
Revised by [3]
3 (bases 1 to 22449)
de Crecy-Lagard, V.
Direct Submission
Submitted (24-MAR-1997) V. de Crecy-Lagard, Institut Pasteur, 25
Rue de Dr. Roux, Paris Cedex 75724, FRANCE
On Mar 25, 1997 this sequence version replaced gi:1483602.
Location/Qualifiers
1. .22449
/organism="Streptomyces pristinaespiralis"
/strain="SP92"
/db_xref="taxon:38300"
/clone="piBV3 cosmid"
115. .7890
/gene="snbc"
115. .7890
/gene="snbc"
/codon_start=1
/transl_table=11
/product="Pristinamycin I synthase 2"
/protein_id="CAA67248.1"
/db_xref="GI:1483603"
/db_xref="SPTREMBL:Q54959"
/translation="MLVPDDARLPLTAALGLWFAQADPHSPVFNTGEVVELTGGLD
ADRFCAALRAVARTDALRLRHTDAEGTQTQSVADDVFTLHRVDVSDDEDPAAAD
AMRDVIARLDVTCGRPLTQALFTLGGRRHWFQHVHLLLDGTFRLVAERVAAD
VRAGAGDAEAGDEPGIRDLAAEADMSADYADQRAHWLATPHPTPTVLAGTTPA
PGAGALRTASAGPEVLGADAVAAEATGQVAFPAAPAAVYVQMLTERTVTLCMP
VTARTSAARRAPGTVANVPALHVPAPAAPLAELVASVRDALEPLLHORYPYEAR
RELGAVERAYGPVVNMPYGRALDGGPAAASVHPVSTGCPVEDLKFSVVEEDGSGAR
IDLEANAAYGSESTDALHARVRELCFAEAGPDTPLGRHIAAPAPAAAPAAPTTGG
PALHTLFAARAAATPHAVASHDGHLYAQDLARANLHRLHLEVOGAPERLVALTL
PSRLDLVALLAVLKSAAAYPLDPDYPADRLAWLLEDAAPVLVDLPQAMTBDLAGY
DTAPTAVDGAHPAYVITSGTGRPKGVVIPHNSVRLFTSTDHWFGRGPDVMTLF
HSYAFDSWEIWLKGLHGGRLVVVYHVSRSPPDFDLDLAREKVTVLNQTPTAFHOL
DAADRARTAAPELALRYVFGGELDARLADYARRGTAARLVNMYGTTETTVHVT
APLPGCHAVPGTPTSLGGPIPLDTLPRVLDALRLPVPPTGFTGELYVAGAGLAGYLNRP
ALTAQRFPADPYGAPCTRMRYRTGDLVRHLLDGTAYLGRDDQVVKIRGRIEELTEN
VLATPHGVAQAAVREDRHGLRLAAAYVPTTGPTEPDVAGWRLAGLQPLHSLPAS
FTLLAALPLTANGKLDRAALPAPAPAPETADEPADPADGEAALRRIFADTGTPTV
GADSDFTLGGHSLLAGLAAVROTLGAEITQDLFEAPTAGLAARVTAATGTRPP
LVPRDHGDGPRPTSVAQHRMNVLREVEGAAPTYNMPALRLSCTLNRLTAALYVDV
VRHESLRTVPRGEGVTPRLLPPEQARPWVIESRVDDQLPALTAANAREPNDQCQ

gene
CDS

Qy	333	CCCACCCGCTCCACGAGAGAAACCAAGAAATAGACCGCGGCATCTCACCGCCGCGCGTG--	390
Db	1131	GGTCCGGAAACGGGCGCTACGGCCCGCTCGTCAACGTCATGCCCTACGGCGCGCCCTGGA	1190
Qy	391	--CTCCAGGAGACCGCGCGCGGAGACATCGCCCTGATCTTCGGCGGTGCTGTGCGCGCT	447
Db	1191	CTTCGGCGGCCCGCGCGCTGCGTGACCCGCGTGTCCACCGCGCCCGTGGAGAGACCTGAA	1250
Qy	448	GCAGCCCGGCATCCGCTCCGGGCTCGCTCCGTGCGGTCTACACATGACGCCGCCACGC	507
Db	1251	GTTCCTCGTGTACGAGAGCCGACGGCTCCGCGCGCGGATCGACTTCGAGGCAACGC	1310
Qy	508	GGTGGCT-----CTGGGACCCCAACGCCCGGTACCAACCCAGCGCGCTACCACT	560
Db	1311	GGCGGTGTACGGCGAGAGACGACGACGCCCATCTGGCGGTCTTCGTCGCTTCCTGGG	1370
Qy	561	GTACTACCTCACTCCGACCAAGAACAGCGCCCGCGGCTGGGACACACGGGAGCAGAC	620
Db	1371	CGCCTTCGCGAGCGCGGCGCGGACACGCGCTCGGCGCGGATCCACGCGCGCGCGCCGC	1430
Qy	621	CGACGGCGTCCGCTTACGACCAACGCGACCGCTGATCGCTGGCGGCCGACCTTCCTCGT	680
Db	1431	CGCCCCCGCGCGCGCGCGCCACACACGCGCGCGCGCGCTGCACACCTGTTCGCGCG	1490
Qy	681	GTGTTCGGGTGCGCGGTCTGTCGGCACGCGCAACACGCGAGGTTTGGCGCGCGCGGCT	740
Db	1491	CCGCGCGCGCCACCCCGACGCGGTGCGCTCAGCCACGACGAGAGACCACTCACCTA	1550
Qy	741	CGTCGCGCTCGCACCCAGCGACGAGCGGTCTCGCAAGTACGAGGACGAGTACCTCTA	800
Db	1551	CGCGCAGCTCGACCGCGCGCGCAACGCGCTCCGCCACCGGTCTGTCGAGCAGGCGCAGG	1610
Qy	801	CTGTCTGACCGACGCGCGGTCTCAGTTACGCCCTCGCCCGCGCGCTATCGTCAACAC	860
Db	1611	CCCCGAACGCTCGTCGCGCTCACCTGCCGCGCTCCCTCGACTGTGTCGCGCTGCT	1670
Qy	861	CGACGGTGGCGCGCACCAACCGCGCGGAGATCGAGAACCGCGAGTGTTCGCGAGCC-	919
Db	1671	CGCGGTCTCAAAATCGGCGCGCGCTACCTGCGCGCTCGACCCCGACTACCGGCGCGACG	1730
Qy	920	-----CCAAGATCCATGGGACACGCCCGCGGGAATGGGTCTGCGTCACTCGGACGACT	974
Db	1731	GCTGGCTTGGATGCTCGAGGACGCGCGCCCGCTCTCTCTCTCGACCGCGAGCGATGAC	1790
Qy	975	GCGGTACCGCGCGCTTACACCTCGCGCAACTCGCGGACCTGGACACTTCGCGCAACTT	1034
Db	1791	CGAGGACCTCGCGGTATCCCGGACACGCCCGCGGTACCGCGGTGACGCGCGCCACCC	1850
Qy	1035	CGACTACCGGAACACGCCCTCGCGGCAATCGAGTGCCTCCCGACCTGTTTCAGATCACCG	1094
Db	1851	CGCCTAGCTCATCTACACCTCGCGTCCACGCGCGCGCCCAAGGGCGTGTGTATCCCGCA	1910
Qy	1095	AGACGAGGGACACGCCACTGGGTGCTGCGCGCGAGATGAGCGCTACGGCATCGCGCT	1154
Db	1911	CAGCAACGTTGTGCGCTTTCACCTCCACCGACCACTGGTTCGCGCTTCGCGCCCGCACGA	1970
Qy	1155	CCCCATGAC---GTACGCTACTGGACAGGACCTTGGGACGCGGAGAGTTCACAGCCGA	1211
Db	1971	CGTGTGGACCGCTTCCACTCTACGCTTCGACTTCTCCGCTTGGGAGATCTGGGCGCG	2030
Qy	1212	CGACCTCACCCCGCAATGGCTCG-----ACTGGGCGCTGGGACTGCTACGCGCG	1259
Db	2031	CTGTCTGACGGCGCGGCTCGTCTGTCGTCGCTACACGTCAGCGCTCGCGCGGGGA	2090
Qy	1260	CGTCACTTGGCCATCGATCGACGCGCGCGAGACCAAGCGCTCCGCCATTCGCGTGATGAA	1319
Db	2091	CTTCTCTGACCTGCTGGCCCGGGAAGGTCACTGCTCTCAACGAGACACCGACCGCTT	2150
Qy	1320	CAACTGGAAGTACGGCGCACGGAGCTCCCAACGACGACATCCCAAGCTACACGGCA	1379
Db	2151	CCACGAGCTCGACGGCGGACGCGCGCGGTACGCGCGCGCGCGGACTGGCCTCAGGTA	2210
Qy	1380	GAACTCGATGTCGCGGAGCTGCGGCTCGCGCGGACAGCGCTGGCGGCTGGTACACCTCT	1439

RESULT 15	
AF210249	
LOCUS	77457 bp DNA linear BCT 29-AUG-2000
DEFINITION	Streptomyces verticillus bleomycin biosynthetic gene cluster, complete cds.
ACCESSION	AF210249 AF149091
VERSION	AF210249.1 GI:9937210
KEYWORDS	
SOURCE	Streptomyces verticillus.
ORGANISM	Streptomyces verticillus
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE	1 (bases 74421 to 74593)
AUTHORS	Du, L., and Shen, B.
TITLE	Identification and characterization of a type II peptidyl carrier protein from the bleomycin producer Streptomyces verticillus ATCC 15003
JOURNAL	Chem. Biol. 6 (8), 507-517 (1999)
MEDLINE	99352421
PUBMED	10421758
REFERENCE	2 (bases 1 to 77457)
AUTHORS	Du, L., Chen, M., Sanchez, C. and Shen, B.
TITLE	An oxidation domain in the BlmII non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the antitumor drug bleomycin in Streptomyces verticillus ATCC15003
JOURNAL	FEMS Microbiol. Lett. 189 (2), 171-175 (2000)
MEDLINE	20389599
PUBMED	10930733
REFERENCE	3 (bases 1 to 77457)
AUTHORS	Du, L., Sanchez, C., Chen, M., Edwards, D.J. and Shen, B.
TITLE	The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a


```
polyketide synthase
Chem. Biol. 7 (8), 623-642 (2000)
20500448
PUBMED
11048953
4 (bases 1 to 77457)
Du, L., Sanchez, C., Chen, M., Edwards, D. J. and Shen, B.
Direct Submission
Submitted (01-DEC-1999) Chemistry Department, University of
California at Davis, One Shields Avenue, Davis, CA 95616, USA
On Aug 29, 2000 this sequence version replaced gi:5326869.
Location/Qualifiers
1..77457
organism="Streptomyces verticillus"
/strain="ATCC15003"
/db_xref="ATCC:15003"
/db_xref="taxon:29309"
complement(1199..2185)
/codon_start=1
/transl_table=11
/product="Syrp-like protein"
/protein_id="AAG02342.1"
/db_xref="GI:9937211"
/transl_table="MTGSLVTLPLGGIIPRGRGELTTGAEDVLDGLGADGDMVRAH
GRLRLERATDGLILLHCLPTDGDGVDFHVSVDGDPDPTERSTPRSVKNIY
TSTETPADQPIPMENESYAAHWPSLTYFFCHTADPTGATADGRAVLDLIPAEVR
RFSQGVYTFTRMDGLSWOEAFQTDGRDGRVCHRAHGOEFSWDGDLVLRHRRP
ATAYDPTGAEVFNQAHLPFSSLDPLDRLQVLTLETGNGLPRDALFADGTPIPDAD
LATVRAAYTRAALALPWREGDIMLVNLRMAHGREPPTGERRVLVAMTSADS"
complement(2182..3918)
/codon_start=1
/transl_table=11
/product="peptide synthetase NRPS12"
/protein_id="AAG02343.1"
/db_xref="GI:9937212"
/transl_table="MTLTLRDAFLDQAARTDAHAVVHGDVTVWYRELELRAGMART
LAAGAGPTLVLRPEPPEPFAALLAVLTGAGYVPLADDDPPDRCHLDDCAA
LLLAHSPSRDGLTTPDEALAPAFDAAPFADGAPVAYIYTSRSSGSRKPKVLEQGA
LGAYLAQARSGTDLGSLHSSDFMAYTSLMGPLVSGGAHVLDLKAIASQTP
PPAASARPSFLKVPSPHLPGLGLLPSDCLPTGQLVIGGEALTGSALGPWRAHDPVT
VNEXPQTGATCCCAVTRPGDAPDPAVPIGRPFAGTRLVLDADGPVAVGGVGL
HIAGDQARGVLRPLTEERFPDPAAGSRMYRTGDLVREPRDGLVLRADGQ
VKVSGYRIPEGEIEAVLKGHAGKDCVAVGEADARLVAYVDPDPSPPGTAPAR
HAALGPYMWPAFVTPFELPLTPNGKLDRLDLPAGDPAGDPTPAETLLCEL
ARALGIPEDADADFLTSGGTSITALKLVAGARRVGIRLELTVLRERTVRIILAAQ
DAASPLAEGVPE"
complement(3915..4733)
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG02344.1"
/db_xref="GI:9937213"
/transl_table="MSKNAHWSRIRTCDAQGVVLAVDFYGTGRQETFRHLCDLTD
PVEVHVPAPDGSSTATGAGHLRWTEGDTVLGAPRVRLVGYCAGGVFASALA
DALVERGHRPRVLPNPSAPGVATLTRDFGLTAGMDLLTDGERAALLAETAIRRA
HADAPLVABRYAALYREGDCLICERLGVDSAFCAELAAVLHSYLAITLALDVPPT
PLWGAVALTSREHQGTDFDVEHGFVVARAELLSSQVVAALTALLREHESR"
complement(4741..6024)
/codon_start=1
/transl_table=11
/product="putative transporter"
/protein_id="AAG02345.1"
/db_xref="GI:9937214"
/transl_table="MTARVDATPTYLAVLAVREARAPLLGSCLARMSFVPLPLALLS
VRDATSFVAGLTSGLSATLTLFAPARAKLIDRRSGSLVRLTVPLVGLAVLIT
LAEEAPTAALLAAVAVGAPFAPPTMRVLIHLHGRQPLHLTAYLDSVTEEV
FTVGLPLAGGLIAVAAFLASMTYVLIAGTACFVLSAATAAPASGEADEDPRPHR
PMALPGMRTIVLISGEGVLVQVVLQVLPFTIADHAGSPGAGGILLSSAGSAYGGLA
YGRJAWSTPVRVRVLTGFTLAVPLCLTASPVPAGAFALLVGLICLAPLFTTAYLL
VNDLVTASGTAPEANTWSPANNNGFAAGSAAGVLLDSRGPTVTVTAFAVAATA
VMTVLRRTLLLGAGHEPAAATPADRTAPAEAE"
complement(6021..6773)
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="AAG02346.1"
/db_xref="GI:9937215"
/transl_table="MDFELTRVNPWFSGCDGRPRVRLCALPYAGTAAVFKDPAAL
PPGVELLTAHLPGRDGRTEPPPTAETAEELCALPPSDLPVTVLHSGMALLGYE
EAARLAARGRAPNLLIAAACRPHPPVDPKASGPTVEAELAAATLRPMDTLARDELM
BAVLPALVADITAGDTHRPRPLDPLPKVYIGADDDGTDWRTTLGWRACRACDCEV
VVLPGCHVFLETDRAAVLTVATDLAEAEVGA"
complement(6784..7797)
/codon_start=1
/transl_table=11
/product="clavaminic acid synthase-like protein"
/protein_id="AAG02347.1"
/db_xref="GI:9937216"
/transl_table="MTDPLRLTVALTGESAEVDDLRLTADVPDSTVGLLHRLRLA
AQELPLRIRAEELTGMRLYDSPRALVVTGVDDEIRGTPPAARPAPDPERTDELELL
LLHAALGEAFGWATQNGRLVHDVLPVPGETAQMGSSSETELLHWEDEAFHPLRCD
YVGLLCRLNRHQAATVGMPLDSRLTTEDRAVLLPEFRLIRPDTSHTPAQNAATGRSA
ERFAAIAEMDDAPERVAVFGDPEDPYLRIDPAYMSPAPGAAARAYDTVTALIEDE
LRHVVDAGSLLLVDNYQAVHGKRPFAAYDGRDRWLKRVNITRDLRSRSARRSATS
LLV"
complement(7859..9214)
/codon_start=1
/transl_table=11
/product="putative transporter"
/protein_id="AAG02348.1"
/db_xref="GI:9937217"
/transl_table="MMKSSRLRDLQGGEDPVVAQESPDQAGPTFCQGGDGLNFAAL
AALLEVFPVRLPHHAGLHVHSSQAAVALPQMVASATTGAGDHQAGRRGQ
VTGVGVREAVGGDELGAEDPHARQLRDLGLLWMAAERLADLNPQTQVYQOGL
RGQVSHDGLGVDLPQGRGLLSLGRRRRGVGTAAHVSQPGQPGPAAAEASCRQ
LVAROODQALLRAVVEGPFORGEDAGVAVETVDPDPVGEVGTGQVAVREGOQ
GSHVDREVPVAVGCGDDGVAGVGLRLAPVSAHVADPAAGHVDGLLAVAREGOQ
QRGGRAGVHCVPDLVQVEDLTHGQDRLLVADLLRPORTGLVDHRRPVMPPLARY
H"
complement(9211..17736)
/codon_start=1
/transl_table=11
/product="peptide synthetase NRPS11-10"
/protein_id="AAG02349.1"
/db_xref="GI:9937218"
/transl_table="NASDALSAHLEVSDCRQRGITVTNDGSELAQNVAAVFRERY
AIAPERTALHKGATGYDELNRRRAELTATLADAGAGSTLVAVLPDPLVATLC
ALLKGAACLPLDPCIAGRLREIMADASPDLVLTTRAVAPATGDPVFLDAPPT
CSAVLPRHSAGTASIAIVLPTTPEDEKSENSVSRDMARYLDDPTAGIPARAAILR
LVAPLLSGRLVLDADETRPRPTREAPDMVEDVVAQVCAVLGVDRVGRVDRDFDL
GGKSLAAVQVVARLKKLLGVELPLRALFDAPTVTEELAAVRAEQGGGVREAEALEP
VGRSEPLPLSFAQQRLWFLDRMLPDRAFYTMCDFAFRVGGIDLGALRALRLVLEHE
TLRTAFVERDGVYQOLVGPADGPGARRVAAAPTRVDSLLEPAEREAEVNLVAAEART
PPRPADGALLRVVVARLADDDHVLVSTHHIVSDAWSVGLVDELGRLYRECVTGDP
ALPPAVQYADFVAVQRAWAGPVQEEHLAYWKRALDCAVSVLRPLPMDPRPRAVOSER
GETVFALPDALVAALEKLGREGATLEMTLGAFQVLLIARHAGEDIVVGVPAAGRT
RTEPLGVFFVNTLPLRAICAPGLSDFDLQVREAAFGAFADHQLPEALVEALP
ERDLGNPLVQVFTQLLPTPAARPLDGLTVEYRVPQVAVSQFDSLBTIKRADDSYR
GLINYCPDLFRMRVLEVGHVLTLLIGAAADGPRPIGELPLSDCAERLRLDDGFKR
DAAYAGCVSPERFAEVARTADARVTCGATLTFEELNDRVLAQALLGAGVTR
TPVAVLRPSRSDSVALLAVMRAGVYVPLDDDFADRTAYILDDTAAVTVITRLPA
LPGRHLVDRPPAADGLVPAPRIDPQAAIYIYTSGSTGAPKGVVRRHRSNHLTSAL
QATFLGHDPYLAGADGVPPGDAKLTTLTAPFTFDSMEQSLSMGLAGHELFTVPEDVR
RPSALVFRVREHRIIDVDTTSSQLELLVSHGLDGEWAPSMVWGVGASPSLWRTL
RDORTRCNLNGPTEAVTADTCHDSDPADVPVIGTPLPHTHVRLDRLRPVGV
AGEIYLGSTGLARGVLPALTAREFADVPDTPGSRLYETGDRARPPDGTLEYLG
RDDQIKLRGRFVEPEGEAVLTHHPAVKEAAVDDAHRLVAYTVLAEAGGAGPTDV
RRFAGKLRPAHMPVSAVVVLEALPLTSGNKDLRAKLPAPAGRGRVRFVAPRDMVE
EVVAGVCAVGLGVDRVGGDDFFELGSSLLVYQVMTIRKLLGVEVLELFDAAVY
BELAARVAAARTGELGRGAAPLPDGSRSGLPLPSFAQRLWYLDQAPDSVSNMCD
AVRVGRPLDALARLALTLVERHETLTAFAVERGVPHOVVSPADPAARAAAEVVR
TEARGTDEAVRDLVAAEARTPFPRADGALMEVAVRLADDDHVLVTVTHIVSDGWS
VDILVDELGRLYREHVTGDPAGLPDVOYADFVQKSWMTGPVREHLYLWKRALD
GAPSVLRPLADHPRPAVQSQRTEFEFFPLPVARLEALCREQVTLFMAIFGAFQV
LLARYSGQDDVVVGVPTANRTARETEPLVGFVNTLPVRVACSPELSFLALDVRVEA
```

ALGAFADHLPPEALVEALAPBRDLGHHPVLQVTFOLLIDAPDERLVLHGTDCVSLGFG
GTVSRDLSDLVVSGRRGRVLTICPDILDFRPRMEVLACHYLDLPAALAAADDPLRVG
DLPSDDRLVLRRCVAAETPVACVLRPSADAVALLAVLRAGVYVVPVDPWPWSEGR
LDTRVTLAARLRRCVAAETPVACVLRPSADAVALLAVLRAGVYVVPVDPWPWSEGR
VAHLDTEAPVYITRDLPAQGRVHLDRPADRDPLRLHRDOAAV11FTSGST
GAPKVVVRHGSILYHLGHVRMAEGGPRNVAHTTAMTFDPSLEOFMLVAGHTLHV
APEVRDRDEALVALVRRAAIDVNTVPSHLLIEAGLLEGRVPGTVLVEGEAVPA
ALWRTLRREAGTREFNLGTEATVDATDCHLSDPADVPVIGTUPPHTHVRVLDRL
RVPVPGAGEIYLGSTGLARGYLNRPALTAQREAVDPYDTPGSRLLRTGDRARWRPD
GTLEYLGRYDDQIKIRGFRVEPEIEAVLTHHPVAKAEAVTATDGAARLVALVMPA
PRPHGSDAGAPAOVEEMNAVFEATHYDAADELTEFKWMSDLSLTPAPAEHMR
EWDVTVARLPERPAERVLEIGSTGLLMWRLLPHVTEYTGTFDESPADVMDLGR
RPAHRVRLHREATDFTGVRASSTDLVVNSVQYPPDRAYLDTVLARALDADGR
VFVQGNRIALAPQFVRQALAHAGAAADVARAAGEAFAMDEGLLYDQVHDLASL
ARSPTVGYEILPRGRHRNEMSLYRVDVLHVGGDRPAAPAEVLTWQVHDLASL
SARLRGGDPDLVRGVANDRLTRONELLDAPRTAVETEDLWGLADSTPYRVSVW
AAADPRGMDVLLVRDADHDDGPLLVPVPEPSAPLTPPTPHPSAROGGSAADGLR
SWLAERLPAHLLPARITEVDALPRTGTGKLDRGALGGLVYAGRGARAGRPATAPRTG
LERTLADARVRLGULPEVGHENFALGGSLLAVRAVARCRRAGVRLTVROLLSEQT
VAALAAALEEESQ*
complement(18307..18660)
/codon_start=1
/trans_table=11

Query Match 5.6%; Score 112; DB 1; Length 77457;

Best Local Similarity 42.4%; Pred. No. 0.00038;
Matches 794; Conservative 0; Mismatches 1065; Indels 13; Gaps 3;

QY 9 CCCCGACTTCCCTCGACGACACCGTCCCTTACCGCGCGACCGCGCGCCGACTGCT 68
DB 38086 CGCCAGGACCGCGCTGACAACTGTCTCACTCGCCCGCGGCTTCCGCGCGTGA 38145
QY 69 CTTAGCTTAGAGGGCCCTCTCTGAGGTCTTCTGTTGGGACGGTGAAGCGACTGCGTC 128
DB 38146 GGAGCGCACCCCGCGCGCTGCTGACCGTGAACCGGTGCCACACGTGTCGGCGA 38205
QY 129 GAACCTGGTCTCTGGGGCGCGGTGTGACCGGAGCCTCGAGACGCGACGGCAGG 188
DB 38206 CGACCTCGCCACCCCGAGACGCCACCGTCCCGGGCGCGGCAAGGTGATCCCCGGGA 38265
QY 189 AACCGTGCAGCTGACCGCATGACGTCGAGGCGCGCCAGCGATGCTGACCGCCCTGAACC 248
DB 38266 GTACCGTGTGATCGCTGCACCGCTTGGACGTGGAGCGGGCTGGACCGCGAGCGGT 38325
QY 249 TGGCGCGGTCTGGGTGAGAGGGTCTCCACCCCGACAGCTCTCTCTTACCGGTGCC 308
DB 38326 GCGGAGCTGATGCTCGGGAATCGCGCGCGCGCGAGACCACTACCGCTCGCG 38385
QY 309 GAACAGGTGGAGCTTCTGCGGCCACCCCTCCACAGAGAGAACCAATGACGCC 368
DB 38386 CGGCGACGCGGCTTACACCCCTCGCCCGTCCGGCAGCCCTCCCGCGCGCACCGGAACG 38445
QY 369 GGCATCTACGCGCGCGCTGCTCCAGGAGCGCGCGCGGAGACACTCGCCCTGATCTT 428
DB 38446 CCGCGCGGTCCGCGCGCGCTTACCTGCTGCGGGCGCTCGGGCGCATCGCCT 38505
QY 429 CCGCGGTGCTGTCGCGCTGACGCGCGGATCCGCTTCGGGTGCTGCTGCTGCCGTCTA 488
DB 38506 CCACTCGCGAGTAC ---TGGCGCGCGCGCGCGACACCGTCTCTACCCACCGCGG 38562
QY 489 CCACATGACGCCCGCGCGGTCTGCGACACCCCAACCGCGGTACACACCGG 548
DB 38563 GCCCTTCCCGCGCGCGGTGGAGCGGTGCGCGGGACACCCGAGCGCGCGT 38622
QY 549 CGCTACCAAGCTTACTTACTCTCCGACCAAGAACAGCGCGCGCGGTGGGACCA 608
DB 38623 CPTCCGCGGCTCGCTCTCCCGCCACCGCGCGCACGGTCTGCTGCGCGCGCGCA 38682
QY 609 CCGGAGACAGACGAGCGGTCTGCTTACAGCACACCGACCGCATGATCGCGTGGGCC 668
DB 38683 CTTACCGACACGAGCGCATGCGCGCTTCGCGGACGAGGTGGAACAGGCCACGCGCC 38742
QY 669 CGACTTCCCGGTGCTGCGGGTCTGCGGACCGCGACCGCGGACCGGTTGCG 728

DB 38743 CGTCCGGGGGTGTGTGACGGGGCGGGGTGCCCGACACCGCCCGCATGATCCAGCGTCG 38802
QY 729 CGCGGGCGCGGTCTGCGCGTTCGCGACCCAGCCAGCCAGCGCGCTCGCGAAATACCAAGA 788
DB 38803 CGACCGAGCGGCGACGACCGCGCGCTTCCGCGCCAAACTGACCGCGCACCTCGTCTGGA 38862
QY 789 GCAGTACTTACTGTGTGACCGACGCGGGTTCACGTTCACCGCCCTCGCCGACCGCGT 848
DB 38863 CGAGGTTCGCCACCGCGACCTCGACTTCTCTGCTGCTCTCGATCGGACCGT 38922
QY 849 CATCGTCAACACCGACCGTTCGCGCGCCACACCGCCCGCGAGATCAGAACCGCGAGTG 908
DB 38923 GCTGCACAAGCTGAAGTTCGCGGAGGTTCGGTACGTGGCGGCAACGAGTTCCTCGACGC 38982
QY 909 GTTCCCGGACCCCAAGATCCACTGGGACACCGCCCGCGGAGAAATGGGTCTCGCTCATCGG 968
DB 38983 CTATGCGCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCTGATCGCTCGCTCGACCGA 39042
QY 969 ACAGTCCGCTAGCCCGCTTCTATACCTCGCCGAACTCGCGGACCTGGGACACTTTCGCGC 1028
DB 39043 CTGCGGGAGTGGGCGATGTGGCGCGCGCGCGCGCGCGCGCTGACCGCGCTACGCGAC 39102
QY 1029 CAATTCGACTACCCGAAACACCGCTTCGCGGATCGAGTGGCCCGACCTGTTTCGAGAT 1088
DB 39103 CGCGCGCGACTGCGCTTACCGCGCGCGCGCGCGCGCTGCTCGCGCGATACGCCCCGAGGA 39162
QY 1089 CACCGAGACGAGGAGCGGACGACCTGGGTGCTCGCGCGCGAGCATGGACGCTTACGGCAT 1148
DB 39163 GGGGCTCGAGCTTCTCCCGCGGTGCTGCGCGCGCGACACCGCGCGGACGTCATCGTCTC 39222
QY 1149 CGGCTCCCCATGAGCTTACGCTTACTGGACAGGACCTGGGACGGCGGAGAGTTCACAGC 1208
DB 39223 GGCCAGGACTCGACGAACTCTCGCGCGCGACGCGCGGTACACACCGACGACCACT 39282
QY 1209 CGAGGACTCACCCCGAATGGGTGCTGAGTGGGCTGGGACTGTGACGCGCGCTGACCTG 1268
DB 39283 CGCGCGCTCGGCGACCTGAGGATCGCGCGCGCGCGCGCGCTTCTCGCGCTCGCGCGCGCG 39342
QY 1269 GCCATCGAGCGCGCGCGAGACCAAGCGCTCGGCCATCGCGTGGTGAACAACTGAA 1328
DB 39343 GTACGGGCGCCCGACACGCGCGCGCGGATCGCGCGCTGATCCGCGACCTGCT 39402
QY 1329 GTAGCGGACGAGCGCTTCCCGACCGAGCATCGGAGCGGTACAAACGGGAGAGACTCGAT 1388
DB 39403 CGCGCTCGAAGACGTCGCGCTCGACGAGCTTCTTCGCGCTCGCGCGGAGACTCGCTGCT 39462
QY 1389 CGTCCGCGAGCTCGCGCTCGCGCGAGCTGGGCGTGTACACCTCTCTGAGCACCCC 1448
DB 39463 CGCCTCGCGCTGCTGTCGAGCTGCGGGAGCGCTACGGGTGGAGATCTTCGTGCGCG 39522
QY 1449 CGTGGGAGCGCTGACGAACTACGTACCGCGCACCCACACACTCCCGCGACCGGACCGTGA 1508
DB 39523 CATGTTGACGAGCCCA ---CGGTGCGCGCTGCGCGCGCGCGCGCGCGCGCG 39575
QY 1509 CGGAGCGCGCTTCTGCGATGGAACGCGGAGCATAGAGATCGAGCTGACATCGCTG 1568
DB 39576 CGGAAGAGACCGCGCGCGGAGAGGTGTGCT ---GTGACACGCGCGCGCGCTCACCGA 39632
QY 1569 GGACCGCGGACGAACTCGGCTCTCGGTGGCGCTTCCCGCGACGGAACCGCGGACAC 1628
DB 39633 CTTGCTACCGAGCTCGCGCGCGGAGGTGACCTTACGGCGGACGGGACCGGCTGCA 39692
QY 1629 GAACATCGGCAAGTACGAGCAGACCTGTATCGTGCAGCGAGGACCTTCCGACCTTCGCGG 1688
DB 39693 CTGCGCGCGCGCGCGCGCGCTTCCAGGAGCTCTTCCGACCATCTCCGCGCGCGCG 39752
QY 1689 GTACTGCTGCGCGCTTACTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1748
DB 39753 CGACGAACTCTCG 39812
QY 1749 GCACCTCGCGCTCTGCTGACACCGAGCGGTGAGGTCTTCTGCTAACCGCGCGGACAC 1808

Db 39813 GCCGCTGTCTTCGCCCGAGGAAACGGCTCTGGCTCTCCACCAAGTTCCACCCCGCAGACAG 39872
QY 1809 CGTGTCTCTCCAGCAGTCCACTTCGCCGAGGGCGGACACGGGAATCTCGCTCTACACCGA 1868
Db 39873 CGCCTACAACATCCCCCTGCACATGCCCTGGCGGGCCCCCTGAACCCGGCGCCCTGGC 39932
QY 1869 CGCGGGCCCCGC 1880
Db 39933 CGCGCCCTGGC 39944

Search completed: May 10, 2003, 08:28:52
Job time : 6087.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 02:34:41 ; Search time 434.5 seconds
(without alignments)
10371.113 Million cell updates/sec

Title: US-09-868-328B-2

Perfect score: 2001

Sequence: 1 gcggtgcaccccgacttccc.....atcgacacgtctctgtgtt 2001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDS2/qcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2001	100.0	2001	22	AAF30918
2	2000	100.0	2000	22	AAF30919
3	945.2	47.2	1752	24	ABL54835
4	931.8	46.6	1551	20	AAI19827
5	912	45.6	1467	24	ABL54834
6	909.6	45.5	1452	20	AAI19826
7	147.6	7.4	65140	22	AAI17184
8	147.6	7.4	125401	22	AAI17186
9	134.8	6.7	114955	20	AAI53491

10	119.8	6.0	3849	22	AAF25795
11	112	5.6	5857	21	AAA58471
12	109.2	5.5	27541	22	AAI17185
c 13	101	5.0	4403765	22*	AAI199683
14	99.8	5.0	114955	20	AAI53491
15	98.6	4.9	77536	21	AAA14651
16	97.6	4.9	2712	20	AAZ06825
17	97.6	4.9	2712	20	AAH74538
c 18	94.8	4.7	4411529	22	AAI199682
19	93	4.6	5392	15	AAQ64201
20	89.8	4.5	30690	21	AAA92301
21	89.8	4.5	30690	22	AAH79277
22	89.8	4.5	31422	21	AAA92302
23	89.8	4.5	31422	22	AAH79278
c 24	89.8	4.5	77536	21	AAA14651
25	89.4	4.5	2291	9	AAH80309
26	88.6	4.4	2271	11	AAQ06844
27	88.6	4.4	3391	22	AAQ88698
28	88.2	4.4	882	24	ABK73444
29	85.6	4.3	44377	18	AAI78508
30	85.6	4.3	44377	18	AAI80414
31	84.4	4.2	2811	21	AAI55794
32	84.4	4.2	18034	21	AAI55841
c 33	84.2	4.2	985	24	ABQ44816
34	84.2	4.2	985	24	ABQ44817
35	84	4.2	4074	23	AAI76405
c 36	84	4.2	8438	15	AAQ73500
37	83.2	4.2	11916	22	AAH79279
38	83.2	4.2	24379	18	AAI93095
39	83.2	4.2	24379	19	AAI25925
40	83	4.1	1908	21	AAI61403
41	83	4.1	4403765	22	AAI199683
42	82.2	4.1	2061	22	AAI86409
43	81.8	4.1	1773	24	ABK86017
44	81.8	4.1	1773	24	ABK86018
c 45	81.4	4.1	1127	21	AAA02477

ALIGNMENTS

RESULT 1

AAF30918
ID AAF30918 standard; DNA; 2001 BP.

XX AC AAF30918;

XX XX 09-JUL-2001 (first entry)

DE Arthrobacter ureafaciens levan fructotransferase DNA.

XX Rhee S, Song K, Kim C, Ryu E, Lee Y;
PI WPI; 2001-308483/32.
XX P-PSDB; AAB82301.
DR Producing difructose dianhydride IV from sucrose, involves reacting,
XX sugar solution in the presence of levansucrase to produce levan, and
PT produce levan solution in the presence of levan fructotransferase to
PT produce DFA IV.
XX
PS Claim 4; [Page 48; 72pp; English.
XX
XX The present sequence is that of *Arthrobacter ureafaciens* K2032 DNA
CC encoding levan fructotransferase (see AAB82301). The DNA was
CC obtained by PCR amplification of genomic DNA using degenerate
CC primers (see AAF30921-22). Expression vector pUDFA81 carrying
CC the levan fructotransferase gene is claimed. Also claimed is
CC *Escherichia coli* JUD81 (KCTC 0877BP), prepared by transforming
CC *E. coli* DH5-alpha with pUDFA81. A claimed process for producing
CC difructose dianhydride IV (DFA IV) from sucrose comprises
CC subjecting sugar solution to reaction at room temperature or lower
CC in acidic buffer of pH 3.0-7.0 in the presence of a levansucrase
CC derived from *Zymomonas mobilis* to produce levan, purifying the
CC levan from the reaction solution, and subjecting it to reaction at
CC 25-30 degree C for 3-10 hours in acidic buffer of pH 3.0-7.0 in
CC the presence of levan fructotransferase, preferably obtained by
CC cultivation of *E. coli* JUD81. The product is useful as a
CC low-calorie sweetener.
XX
XX Sequence, 2001 BP; 337 A; 813 C; 588 G; 263 T; 0 other;
SQ

Query Match 100.0%; Score 2001; DB 22; Length 2001;
Best Local Similarity 100.0%; Pred. No. 2e-310;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGGTGACCCGACCTCCCTCGAGACACACCGTCCCTACCGGCGACGCCCGGCC 60
DB 1 GCGGTGACCCGACCTCCCTCGAGGACACACCGTCCCTACCGGCGACGCCCGGCC 60
QY 61 CGACTGCTCCTCAGCTAGACGGGCGCCCTCTCGAGGTCTTCGTCGGGACGGTGAGGG 120
DB 61 CGACTGCTCCTCAGCTAGACGGGCGCCCTCTCGAGGTCTTCGTCGGGACGGTGAGGG 120
QY 121 ACTGCTGGAACCTGCTCTCTGCGGGGCGCGGCTGTGACCGCGAGCCTCGAGACGGCA 180
DB 121 ACTGCTGGAACCTGCTCTCTGCGGGGCGCGGCTGTGACCGCGAGCCTCGAGACGGCA 180
QY 181 CGGCCAGGAACCGTGACGTCGACCGGATCGAGGTCTGAGGCGCCCGACGATGCTGACGCC 240
DB 181 CGGCCAGGAACCGTGACGTCGACCGGATCGAGGTCTGAGGCGCCCGACGATGCTGACGCC 240
QY 241 CTGGAACCTGCGCCGCTTCTGGGCTCAGCAGGCTCCACCGGACGAGCTCTCTTCTAC 300
DB 241 CTGGAACCTGCGCCGCTTCTGGGCTCAGCAGGCTCCACCGGACGAGCTCTCTTCTAC 300
QY 301 CGCTGCCCCGAACAGGGTGGAGCTTCTGTCGCGCCACCCGCTCCACGAGGAACAGCA 360
DB 301 CGCTGCCCCGAACAGGGTGGAGCTTCTGTCGCGCCACCCGCTCCACGAGGAACAGCA 360
QY 361 ATGACCGCGGCATCTACGCGCGCGCGCTGCTCCAGGAGCGCGCGGACGACTCGCC 420
DB 361 ATGACCGCGGCATCTACGCGCGCGCGCTGCTCCAGGAGCGCGCGGACGACTCGCC 420
QY 421 CTGATCTTCGGGGGTGCTGTGCGGCTGACGCGCGGGCATCCGCTCGGGGCTCGCTCCGT 480
DB 421 CTGATCTTCGGGGGTGCTGTGCGGCTGACGCGCGGGCATCCGCTCGGGGCTCGCTCCGT 480
QY 481 GCGGTCTACCATGACGCCCCCGAGCGGCTGGCTGTGCGACCCCGCAACCGCGGTACCC 540
DB 481 GCGGTCTACCATGACGCCCCCGAGCGGCTGGCTGTGCGACCCCGCAACCGCGGTACCC 540
QY 541 ACCGACGGCGCTACGAGCTGTACTACCTGCACTCCGACCAACACGCGCCCGCGGCC 600

DB 541 ACCGACGGCGCTACGAGCTGTACTACCTGCACTCGACCAACACGCGCCCGCGGC 600
QY 601 TGGGACACGCGAGCAGCAGCCGCGCTTACGCAACACGCGACCGCTGATGCCG 660
DB 601 TGGGACACGCGAGCAGCAGCCGCGCTTACGCAACACGCGACCGCTGATGCCG 660
QY 661 CTCGCGCGCAGCTTCCCGGTGCTGCGGGTCTGCGGACCGCGGACACGGAAC 720
DB 661 CTCGCGCGCAGCTTCCCGGTGCTGCGGGTCTGCGGACCGCGGACACGGAAC 720
QY 721 GGGTTGCGCGCGCGGTGCTGCGGCTCGGACCGCGGACCGCGGCTCCGCAAG 780
DB 721 GGGTTGCGCGCGCGGTGCTGCGGCTCGGACCGCGGACCGCGGCTCCGCAAG 780
QY 781 TACCAGGACGAGTACCTCTACTGCTGACGCGCGGGGTTCACGTTACCGCGCTGCC 840
DB 781 TACCAGGACGAGTACCTCTACTGCTGACGCGCGGGGTTCACGTTACCGCGCTGCC 840
QY 841 GACCCGCTCATGCTCAACACCGAGCTGCGGCGCGCCACACCGCGCGGAGATCGAGA 900
DB 841 GACCCGCTCATGCTCAACACCGAGCTGCGGCGCGCCACACCGCGCGGAGATCGAGA 900
QY 901 GCGGAGTGGTTCGCGACCGCCAGATCCACTGGGACACCGCGCGCGGAGATGGTCTGC 960
DB 901 GCGGAGTGGTTCGCGACCGCCAGATCCACTGGGACACCGCGCGCGGAGATGGTCTGC 960
QY 961 GTCATCGGACGAGCTGCGGTACGCGCGGTCTACACCTCGCGGAACTGCGCGACTGGA 1020
DB 961 GTCATCGGACGAGCTGCGGTACGCGCGGTCTACACCTCGCGGAACTGCGCGACTGGA 1020
QY 1021 CTTGCGCGCAACTTCGACTACCGGAACACCGCTCGGGGCACTGAGTGCCTCGGACCTG 1080
DB 1021 CTTGCGCGCAACTTCGACTACCGGAACACCGCTCGGGGCACTGAGTGCCTCGGACCTG 1080
QY 1081 TTCGAGATCACCGACAGCGGACACCGCTGCGGCTGCTCGCGCGGAGATGGAGGCC 1140
DB 1081 TTCGAGATCACCGACAGCGGACACCGCTGCGGCTGCTCGCGCGGAGATGGAGGCC 1140
QY 1141 TAGGGATCGGCTCCCATGAGTACGCTACTGACAGGACCTGGGACGCGGAGCAG 1200
DB 1141 TAGGGATCGGCTCCCATGAGTACGCTACTGACAGGACCTGGGACGCGGAGCAG 1200
QY 1201 TTCACGCGCAGACCTACCGCGCAATGGCTGCGAGTGGGGTGGGACTGGTACGCGGCC 1260
DB 1201 TTCACGCGCAGACCTACCGCGCAATGGCTGCGAGTGGGGTGGGACTGGTACGCGGCC 1260
QY 1261 GTCACCTGGCCATCGATCGAGCGCGCGGACCAAGCCTCGCCATCGCTGGATGAAC 1320
DB 1261 GTCACCTGGCCATCGATCGAGCGCGCGGACCAAGCCTCGCCATCGCTGGATGAAC 1320
QY 1321 AACTGGAAGTAGCGCGACGCGGACCTCCCGACCGAGCATCCGACGGCTACACGCGGAC 1380
DB 1321 AACTGGAAGTAGCGCGACGCGGAGCTCCCGACCGAGCATCCGACGGCTACACGCGGAC 1380
QY 1381 AACTCGATCGTCGCGAGCTGCGGCTGCGCGGACGCGCTGGGGTGGGACTGGTACCCCTCCTG 1440
DB 1381 AACTCGATCGTCGCGAGCTGCGGCTGCGCGGACGCGCTGGGGTGGGACTGGTACCCCTCCTG 1440
QY 1441 AGCACCCGCTGCGAGCTGAGAACTACGTCACCGCCACCACTCCCGGACCGG 1500
DB 1441 AGCACCCGCTGCGAGCTGAGAACTACGTCACCGCCACCACTCCCGGACCGG 1500
QY 1501 ACCGTCGAGCGGCGCTCCTGCAATGGAACGCGGAGCATACGAGATCGAGCTCGAC 1560
DB 1501 ACCGTCGAGCGGCGCTCCTGCAATGGAACGCGGAGCATACGAGATCGAGCTCGAC 1560
QY 1561 ATCGCTGGGACACCGCGAGCAGCTGCGGATCTCGGTGGGCGGCTCCCGGACGGAAC 1620
DB 1561 ATCGCTGGGACACCGCGAGCAGCTGCGGATCTCGGTGGGCGGCTCCCGGACGGAAC 1620
QY 1621 CGGACACGGAACATCGGCAAGTACGAGGAGCAGCTGTACGTCGCGGAGGACCTCCGAC 1680

Db 1621 CGGCACAGCATCGGCAAGTACGAGGACAGACCTGTACGTGACACGAGACCCCTCCGAC 1680
QY 1681 CTCGGCGGTACTCGTCCGCCCTTACTCGCGAGCGCGCCGCCATCGACCCCGCGCC 1740
Db 1681 CTCGGCGGTACTCGTCCGCCCTTACTCGCGAGCGCGCCGCCATCGACCCCGCGCC 1740
QY 1741 CGATCGGTGACCTCGGCGATCTCGTGCACACCCAGAGCGTTCGTCCTCAACGCC 1800
Db 1741 CGATCGGTGACCTCGGCGATCTCGTGCACACCCAGAGCGTTCGTCCTCAACGCC 1800
QY 1801 GGCACACCGTCTCTCCAGCAGTCCACTTCGCGAGGCGGACACGGAATCTCGCTC 1860
Db 1801 GGCACACCGTCTCTCCAGCAGTCCACTTCGCGAGGCGGACACGGAATCTCGCTC 1860
QY 1861 TACACCGAGCGGCGCGCCGACACTTCACCGCATCTCGTCCGCGAGATTGGCCAGGCG 1920
Db 1861 TACACCGAGCGGCGCGCCGACACTTCACCGCATCTCGTCCGCGAGATTGGCCAGGCG 1920
QY 1921 ATCTAGGCGATGCACACACCGTTCACCGAAGCGCGCCGCCGAGACGACGCGCGAC 1980
Db 1921 ATCTAGGCGATGCACACACCGTTCACCGAAGCGCGCGCCGCCGAGACGACGCGCGAC 1980
QY 1981 AATCGACACGTCCTCGTCGTT 2001
Db 1981 AATCGACACGTCCTCGTCGTT 2001

RESULT 2
AAAF30919
ID AAF30919 standard; DNA; 2000 BP.
XX AAF30919;
XX
XX
XX 09-JUL-2001 (first entry)
XX
XX Arthrobacter ureafaciens levan fructotransferase DNA.
XX
XX Levan fructotransferase; difructose dianhydride IV; sweetener; ds.
XX
XX Arthrobacter ureafaciens.
XX
XX Key Location/Qualifiers
XX CDS 360..1925
XX FT /*tag= a
XX FT sig_peptide 360..458
XX FT /*tag= b
XX FT mat_peptide 459..1922
XX FT /*tag= b
XX
XX WO200129185-A1.
XX
XX 26-APR-2001.
XX
XX 19-OCT-2000; 2000WO-KR01183.
XX
XX 19-OCT-1999; 99KR-0045302.
XX
XX (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
XX (REAL-) REALBIOTECH LTD.
XX
XX Rhee S, Song K, Kim C, Ryu E, Lee Y;
XX WPI; 2001-308483/32.
XX
XX Producing difructose dianhydride IV from sucrose, involves reacting
XX sugar solution in the presence of levansucrase to produce levan, and
XX reacting levan solution in the presence of levan fructotransferase to
XX produce DFA IV -
XX
XX Disclosure; Page 8-9; 72pp; English.
XX
XX The present sequence is that of Arthrobacter ureafaciens K2032 DNA
XX encoding levan fructotransferase (see AAB82301). The DNA was
XX

CC obtained by PCR amplification of genomic DNA using degenerate
CC primers (see AAF30921-22). Expression vector pUDFA81 carrying
CC the levan fructotransferase gene is claimed. Also claimed is
CC Escherichia coli JUD81 (KCTC 0877BP), prepared by transforming
CC E. coli DH5-alpha with pUDFA81. A claimed process for producing
CC difructose dianhydride IV (DFA IV) from sucrose comprises
CC subjecting sugar solution to reaction at room temperature or lower
CC in acidic buffer of pH 3.0-7.0 in the presence of a levansucrase
CC derived from Zymomonas mobilis to produce levan, purifying the
CC levan from the reaction solution, and subjecting it to reaction at
CC 25-30 degree C for 3-10 hours in acidic buffer of pH 3.0-7.0 in
CC the presence of levan fructotransferase, preferably obtained by
CC cultivation of E. coli JUD81. The product is useful as a
XX low-calorie sweetener.
XX
SQ Sequence 2000 BP; 337 A; 813 C; 587 G; 263 T; 0 other;

Query Match 100.0%; Score 2000; DB 22; Length 2000;
Best Local Similarity 100.0%; Pred. No. 2.9e-310;
Matches 2000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CGGTGACCCCGACTTCCCTCGACGACACCGTCCCTACCGGCGGACCGCCGCC 61
Db 1 CGGTGACCCCGACTTCCCTCGACGACACCGTCCCTACCGGCGGACCGCCGCC 60
QY 62 GACTGCTCTCAGCCTAGACGGGCCCCCTCTCTGAGGTCTTCTGCGGGACGTGAGGCGA 121
Db 61 GACTGCTCTCAGCCTAGACGGGCCCCCTCTCTGAGGTCTTCTGCGGGACGTGAGGCGA 120
QY 122 CTGCGTCGAACCTGCTCTCTGGGGCGGCGGTGTGACCGGAGGACCTTCGAGACGGCAC 181
Db 121 CTGCGTCGAACCTGCTCTCTGGGGCGGCGGTGTGACCGGAGGACCTTCGAGACGGCAC 180
QY 182 GGCCAGGAACCGTGCACGTGACCGGATCGATCGAGGCGCCGACGATGCTGAGCGCC 241
Db 181 GGCCAGGAACCGTGCACGTGACCGGATCGATCGAGGCGCCGACGATGCTGAGCGCC 240
QY 242 CTGAACCTGCCCGGTTCTGGGTGACGAGCGCTCCACCCCGACAGCTCTCTCTTATCC 301
Db 241 CTGAACCTGCCCGGTTCTGGGTGACGAGCGCTCCACCCCGACAGCTCTCTCTTATCC 300
QY 302 GCTGCCCGGAACAGGTGACGCTTCGTGCGGCCACCCGTCACGAGAGAACAGCA 361
Db 301 GCTGCCCGGAACAGGTGACGCTTCGTGCGGCCACCCGTCACGAGAGAACAGCA 360
QY 362 TGAGCGCGGCCATCTCACGCGCGCGTCTCAGGAGCGCGCGCGGACACTCGGCC 421
Db 361 TGAGCGCGGCCATCTCACGCGCGCGTCTCAGGAGCGCGCGCGGAGCACTCGGCC 420
QY 422 TGATCTTCGGCGGTGCTGTGCGGCTGACGCGCGGCGATCCGCTCCGGGCTCGCTCCGTG 481
Db 421 TGATCTTCGGCGGTGCTGTGCGGCTGACGCGCGGCGATCCGCTCCGGGCTCGCTCCGTG 480
QY 482 CCGTCTACACATGACGCGCGCGCGGTGCTGCGACCCCGCCAGCGGTCACCA 541
Db 481 CCGTCTACACATGACGCGCGCGCGGTGCTGCGACCCCGCCAGCGGTCACCA 540
QY 542 CCCAGCGCGCTTACGAGCTGTACTACTCTGACCTCCGACACAGAACAGCGCGCGCT 601
Db 541 CCCAGCGCGCTTACGAGCTGTACTACTCTGACACAGAACAGCGCGCGCGCT 600
QY 602 GGGACACCGGACGACGACCGCGGTGCGCTTACGACACCGACCGGATGATGCCGC 661
Db 601 GGGACACCGGACGACGACCGCGGTGCGCTTACGACACCGACCGGATGATGCCGC 660
QY 662 TCGGGCGCGACTTCCCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 721
Db 661 TCGGGCGCGACTTCCCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 720
QY 722 GCTTCGGCGCGCGCGGTGCTGCGGCTCGGACCCAGCGCGGACCGGATGATGCCGC 781
Db 721 GCTTCGGCGCGCGCGGTGCTGCGGCTCGGACCCAGCGCGGATGATGCCGC 780

```
QY 782 ACCAGGAGCAGTACCTCTACTGTCGACCGCGGGGTTACGTTACACCGCCCTGCCCG 841
Db 781 ACCAGGAGCAGTACCTCTACTGTCGACCGCGGGGTTACGTTACACCGCCCTGCCCG 840
QY 842 ACCCGTCATCGTCAACACACGAGTTCGCGCGCCACACCGCCCGGAGATCGAGAAGC 901
Db 841 ACCCGTCATCGTCAACACACGAGTTCGCGCGCCACACCGCCCGGAGATCGAGAAGC 900
QY 902 CCGAGTGGTTCCCGGACCCCAAGATCCACTGGGACACCGCCCGGAGAAATGGTCTGCG 961
Db 901 CCGAGTGGTTCCCGGACCCCAAGATCCACTGGGACACCGCCCGGAGAAATGGTCTGCG 960
QY 962 TCATCGGACGACTGCGGTAGCCGCTTCTACACCTCGCGCAACCTGCGGACTGGACAC 1021
Db 961 TCATCGGACGACTGCGGTAGCCGCTTCTACACCTCGCGCAACCTGCGGACTGGACAC 1020
QY 1022 TTCGCGCGCAACTTCGAGTACCCGCAACCGCCCTCGCGGCGCATCGAGTGCCTGCT 1081
Db 1021 TTCGCGCGCAACTTCGAGTACCCGCAACCGCCCTCGCGGCGCATCGAGTGCCTGCT 1080
QY 1082 TCGAGATCACCGGACGAGCGGACACCGCCACTGGGTGCTTCGCGCGCAGCATGGAGCGCT 1141
Db 1081 TCGAGATCACCGGACGAGCGGACACCGCCACTGGGTGCTTCGCGCGCAGCATGGAGCGCT 1140
QY 1142 ACGGCATCGGCCCTCCCGATGACGCTACGCTTACTGGACAGGCACTGGGACGCGGAGCAGT 1201
Db 1141 ACGGCATCGGCCCTCCCGATGACGCTTACTGGACAGGCACTGGGACGCGGAGCAGT 1200
QY 1202 TCCACGCGGACGAGCTACACCCGCAATGGCTCGACTGGGGCTGGAGTGTACGCGGCGG 1261
Db 1201 TCCACGCGGACGAGCTACACCCGCAATGGCTCGACTGGGGCTGGAGTGTACGCGGCGG 1260
QY 1262 TCACCTGGCCATCGATCGAGCGCCGCGAGACCAAGCGCCTCGCCATCGCGTGGATGAACA 1321
Db 1261 TCACCTGGCCATCGATCGAGCGCCGCGAGACCAAGCGCCTCGCCATCGCGTGGATGAACA 1320
QY 1322 ACTGGAGTACGCGGACGCGGAGCTGCCACCGACGAGTCCGAGCGGTACACGCGGACGA 1381
Db 1321 ACTGGAGTACGCGGACGCGGAGCTGCCACCGACGAGTCCGAGCGGTACACGCGGACGA 1380
QY 1382 ACTCGATCGTCCGCGAGCTGCGGCTCGCGGACGAGCTGGGGCTGTACACCTCCTGA 1441
Db 1381 ACTCGATCGTCCGCGAGCTGCGGCTCGCGGACGAGCTGGGGCTGTACACCTCCTGA 1440
QY 1442 GCACCCCGTGGCAGCGCTGACGAACCTACGTACCCGCCACACACACTCCCGGACCGGA 1501
Db 1441 GCACCCCGTGGCAGCGCTGACGAACCTACGTACCCGCCACACACACTCCCGGACCGGA 1500
QY 1502 CCGTCGACGCGCGCGCTCTGCCATGGACGCGCGCATACGAGATCGAGCTCGACA 1561
Db 1501 CCGTCGACGCGCGCGCTCTGCCATGGACGCGCGCATACGAGATCGAGCTCGACA 1560
QY 1562 TCGCCTGGGACACCGCGAGCAAGCTCGGATCTCGGTGGCGGCTCCCGGACGGAACCC 1621
Db 1561 TCGCCTGGGACACCGCGAGCAAGCTCGGATCTCGGTGGCGGCTCCCGGACGGAACCC 1620
QY 1622 GGCACAGCAACATCGGCAAGTACGGAGCAGACCTGTACGTCGACCGAGGACCCCTCCGACC 1681
Db 1621 GGCACAGCAACATCGGCAAGTACGGAGCAGACCTGTACGTCGACCGAGGACCCCTCCGACC 1680
QY 1682 TCGCGGGTACTCGCTCGCCCTTACTCGGAGCGCGCGCCCTCCATCGACCCCGGCGCC 1741
Db 1681 TCGCGGGTACTCGCTCGCCCTTACTCGGAGCGCGCGCCCTCCATCGACCCCGGCGCC 1740
QY 1742 GATCGGTGACCTGCGCATCTCTCGTGACACCGACGAGCGTTCGAGGTCTTCGTCACGCGG 1801
Db 1741 GATCGGTGACCTGCGCATCTCTCGTGACACCGACGAGCGTTCGAGGTCTTCGTCACGCGG 1800
QY 1802 GGCACAGCGTGTCTCCCGAGCAGGTCCACTTCGCGGAGGCGGACACGGAATCTCGCTCT 1861
Db 1801 GGCACAGCGTGTCTCCCGAGCAGGTCCACTTCGCGGAGGCGGACACGGAATCTCGCTCT 1860
QY 1862 ACACCGACGCGCGCGCCGCGACACTTACCGGGATCGTCTGCGCGGAGATTGGCCAGCGA 1921
```

```
Db 1861 ACACCGACGCGCGCGCCGCGACACTTACCGGCGATCGTCTCGCGGAGATTGGCCAGCGA 1920
QY 1922 TCTAGCGGATGTCACACACACACCGCTCACCAGCGCGCGCCCGGAGACACGCGCGGACA 1981
Db 1921 TCTAGCGGATGTCACACACACACCGCTCACCAGCGCGCGCGCCCGGAGACACGCGCGGACA 1980
QY 1982 ATCGACACGCTCTCTCGTCTGTT 2001
Db 1981 ATCGACACGCTCTCTCGTCTGTT 2000

RESULT 3
ABL54835
ID ABL54835 standard; DNA; 1752 BP.
XX
AC ABL54835;
XX
XX 01-JUL-2002 (first entry)
XX
DE Levan fructotransferase related DNA sequence.
XX
KW Levan fructotransferase; LFTase; difructose dianhydride; DFA IV;
KW levan; gene; ds.
XX
OS Arthrobacter nicotinovorans.
XX
FH Key
CDS 4..1461
FT /*tag= a
FT /*product= "levan fructotransferase related protein"
XX
XX JP2002017366-A.
XX
XX 22-JAN-2002.
XX
XX 06-JUL-2000; 2000JP-0205756.
XX
XX 06-JUL-2000; 2000JP-0205756.
XX
PA (NIPT ) NIPPON TENSEI SEITO KK.
PA (KAGA-) KAKAKU GIJUTSU SHINKO JIGYODAN.
XX
XX WPI; 2002-287313/33.
XX
XX A new levan fructotransferase -
XX
XX Example 1; Fig 4; 15pp; Japanese.
XX
CC The invention relates to a new levan fructotransferase (LFTase). LFTase
CC decomposes the beta-2,6-fructoside bond of levan in polyfructan to
CC produce difructose dianhydride IV (DFA IV). The LFTase of the invention
CC is used for the preparation of DFA IV. The current sequence represents
CC a levan fructotransferase related DNA sequence.
XX
SQ Sequence 1752 BP; 351 A; 592 C; 506 G; 303 T; 0 other;

Query Match 47.2%; Score 945.2; DB 24; Length 1752;
Best Local Similarity 74.3%; Pred. No. 2.9e-142;
Matches 1237; Conservative 0; Mismatches 413; Indels 14; Gaps 3;

QY 314 AGGTGTGAGCGTTCGTCGCGCCACCCGTCACGAGAGAACCC--AGCAATGACGCGCGC 371
Db 42 AGGTGTGAGCGTTCGTCGCGCCGCGCGCAACTGAGAGGAAACGAATCGATGACGTATGA 101
QY 372 CATCTCACCGCGCGCGTCTCCAGGAGCGCGCGGAGCACTCGCCTGATCTTCGG 431
Db 102 CATCTCACCGCGCGCGTCTCCAGGAGCGCGCGGAGTTCGCTTTCATGAG 161
QY 432 CGGTGCTGTGCGCGCTTCGACGCGCGGCGCATCGCTCGCGGCTCGCTCCGCTCTACCA 491
Db 162 CAATGCCATTCCGTCGCGCGCGCA-----CGCCAGGCGATCCCTCCGCGCATCTACCA 215
```

```
Qy 492 CATGACGCCCCAGCGGTGGCTCTCGACAGCCCCCAACGCGCGGTACACACCCAGCGGC 551
Db 216 CATGACCCCGGCTCGGCTGGCTATGTATCGCGAGGACCCGTATACATAAAGCGGC 275
Qy 552 CTACGAGTGTACTACCTGCACTCGGACAGACAGAACAGCGCCCGGCTGGACACGC 611
Db 276 CTACGAGTGTACTACCTGCACTCGGACAGACAGAACAGCGCCCGGCTGGACACGC 335
Qy 612 GAGCACAGCCAGCGGTGCGCTTTCACGCACACAGCGCACCGTGTATGCGCGTCCGCCGA 671
Db 336 GACCACCGCGAGCGAGTGTCTTACACCCACCATGAGTGGTGATGCCAATGCACCGGA 395
Qy 672 CTTCCCGTGTGGTCCGGTGGCGGTGCGGACCGGACCGGACAGCGAGGTTCCGGCG 731
Db 396 CTTCCCGTGTGGTCCGGTGGCGGTGCGGACCGGACCGGACCGGCGGTTCCGGCG 455
Qy 732 CGGCGCGTGTGCGGCTGCGGACCGGACCGGACCGGCGTCCGCAAGTACCAGGACA 791
Db 456 CGGCGAGTGTGCGGCTGCGGACCGGACCGGACCGGCGTCCGCAAGTACCAGGACA 509
Qy 792 GTACCTTCTACTGCGGAGCGGCGGTTCACGTTTACCGGCTTCCCGGACCGCGTGTAT 851
Db 510 GTACCTTCTACTGCGGAGCGGCGGTTCACGTTTACCGGCTTCCCGGACCGCGTGTAT 569
Qy 852 CGTCAACACCGAGGTGCGGCGGCGGACCGGCGGCGGAGATCGAGAACCGCGAGTGGTT 911
Db 570 TGTGAACACTGATGAGCGGAGCGGCGGACCGGCGGCGGAGTGGAGAACCGGAGATGGTT 629
Qy 912 CCGGACCCCAAGATCCACTGGGACCGGCGGCGGAGAGTGGTCTCGTCTATCGGAG 971
Db 630 CCGGACCCCAAGATCCACTGGGACCGGCGGCGGAGAGTGGTCTCGTCTATCGGAG 689
Qy 972 ACTGGGTACGCGCGGTCTTACACCTCGCGCAACCTGCGGCACTGGACACTTCGCGCGAA 1031
Db 690 GCGCGCTACGCGGTCTTACACCTCGCGCAACCTGCGGCACTGGACACTTCGCGCGAA 749
Qy 1032 CTTGACTACCGGAACACCGGCTCGGCGGATCGAGTGGCGGACCTGTTTCGAGATCAC 1091
Db 750 CTTGACTACCGGAACACCGGCTCGGCGGATCGAGTGGCGGACCTGTTTCGAGATCAC 809
Qy 1092 CGGACGCGGACGCGGCGGCTGCGGCGGAGTGGACGCTACGCGGATCGG 1151
Db 810 CGGACGCGGACGCGGCGGCTGCGGCGGAGTGGACGCTACGCGGATCGG 869
Qy 1152 CTTCCCATGAGTACGCTTACTGACAGGACCTGCGGCGGAGGAGTTCACGCGGA 1211
Db 870 CTTGCCCATGAGTACGCTTACTGACAGGTTTACGAGGAGGAGTTCACGCGGA 929
Qy 1212 CGACTACCGCGCAATGGCTGCACTGGGCTGGGACTGGTACGCGGCGGTCACCTGGCC 1271
Db 930 CAACCTACACACAGTGGCTTGAATGGGATGGGACTGGTACGCGGCGGTCACCTGGCC 989
Qy 1272 ATCGATCGACGCGCGGAGACCAAGCGCTCGCCATCGGTTGGATGACAACTGGAAGTA 1331
Db 990 GCGCGTGGAGACCTTGAGACCAAGCGGCTTGGACAGCGTGGATGAACAACTGGAAGTA 1049
Qy 1332 CGCGCAGCGGCTCCCGACCGCATCGAGCGGTACAAAGGCGGAGTTCATCGT 1391
Db 1050 TGCGCGCGCAAGCTGCGGACCGGCGTCCGATGGCTTACGGGCAAAATTCATCAAC 1109
Qy 1392 CCGGAGCTGCGGCTCGCGGACAGCTGCGGCTGGTGTATACACCTCTCTGAGCACCGCGT 1451
Db 1110 GCGGAGCTGAGGCTGAGGCGCAATCGGCGGCTGGTGTATACCTTGTCTCAGCACCGCGT 1169
Qy 1452 GGCAGCGTGCAGAACTAGTCAACGCGGACCAACACACTCCCGGACCGGCGTCCGAGG 1511
Db 1170 TCCGCGCGTTCGAACTATGCACTCCAGCACCAACCGCTTCGGGACCGCAGTCAACGG 1229
Qy 1512 CAGCGCGCTCTGCGTACGAGCGGATACGAGATCGAGTTCGAGTTCGCTGCGTGGGA 1571
Db 1230 CAGTTTCTGCTGCGTGGAGCGGCGGCGGCTGATGAACCTGGAATCTGATGGGA 1289
Qy 1572 CACCGGACGAAAGCTGCGGCTCTCGTGGGCGGCTCCCGGACGGAACCGCGGACACGAA 1631
```

```
Db 1290 CACGCGACGCAACGTTGGAGTCTCGGTGGCGGCTCGTCCGATGGCAGCGCCATACGAA 1349
Qy 1632 CATCGGCAAGTACGAGCAGACCTGTACGTGACCGAGGACCTCCGACCTCGCGGGTA 1691
Db 1350 CATCGGCAAAATACGTTGAGTGTACGTGATCGGATCGGATCGGAGCAAGCGGTA 1409
Qy 1692 CTCGCTCGCCCTACTCTCGGAGCGCGCCCGGATCGACCGCGCGGCGGATCGTGA 1751
Db 1410 TCGCTGCGACCTACACCGCGCGCGCCCATCGATGCGAACGCGAGATCGGTCCA 1469
Qy 1752 CTTGGGATCTCTGTCGACACCGGAGGCTTCGTCAACGCGGCGGCGGACACCGT 1811
Db 1470 CTTGGGATCTCTGTCGACACCGGAGGCTTCGTCAACGCGGCGGCGGACACCGT 1529
Qy 1812 GCTCTCCAGCAGGTCCACTTCGCGGAGGCGGACACGGAATCTCGTCTACACGACGG 1871
Db 1530 GGTTCGACGAGGTGCACTTCGCGGCGGCGGACACGCGGATCTCCCTCTATGCGGACGG 1589
Qy 1872 CGGCGCGGACACTTCACCGGATCGTCTCGCGGAGATTTGGCCAGGCGATCTAGGCGAT 1931
Db 1590 CGGTCCGCGCAACTTCACCGGATCACCATCGCGAGTTCGGGAACCGCATCTAAGCGTG 1649
Qy 1932 GCACACACACCGGCTACCGGAGCGCGCGCGGCGGAGGACGG 1975
Db 1650 CTTCCCGCGGTGGAAGGACGACGCGGCGGCGTGCAGCAGCGG 1693
```

RESULT 4

AA19827
ID AA19827 standard; DNA; 1551 BP.

XX AC AA19827;

XX DT 10-JUN-1999 (first entry)

XX DE Arthrobacter nicotinovorans levan fructotransferase encoding DNA #2.

XX KW Arthrobacter nicotinovorans; levan fructotransferase; ds.

XX OS Arthrobacter nicotinovorans.

XX PN JP11069978-A.

XX PD 16-MAR-1999.

XX PF 28-AUG-1997; 97JP-0232421.

XX PR 28-AUG-1997; 97JP-0232421.

XX PA (NIOC) NIPPON OIL CO LTD.

XX DR WPI; 1999-247463/21.

XX DR P-PSDB; AAY04105.

XX PT Levan fructotransferase gene - for recombinant production of levan fructotransferase

XX PS Claim 4; Page 10-11; 14pp; Japanese.

XX CC The present sequence encodes Arthrobacter nicotinovorans levan fructotransferase. The present invention also describes a method for the preparation of levan fructotransferase in which a transformant is cultured in a medium and levan fructotransferase is collected from the culture. The method can prepare levan fructotransferase in a large amount.

XX SQ Sequence 1551 BP; 315 A; 524 C; 443 G; 269 T; 0 other;

Query Match 46.6%; Score 931.8; DB 20; Length 1551;

Best Local Similarity 75.8%; Pred. No. 4e-140; Mismatches 367; Indels 12; Gaps 2;

Matches 1184; Conservative 0;

QY 361 ATGACGCGGCGCATCTACGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 420
Db 1 ATGACGCGGCGCATCTCTCGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 60
QY 421 CTGATCTTGGCGGCTGTGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 480
Db 61 CTGATCTTGGCGGCTGTGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 114
QY 481 GCGGCTTACCATGACGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 540
Db 115 GCGATCTACCATGACGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 174
QY 541 ACCGAGCGGCGGTCTGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 600
Db 175 ACAAGCGGCGGTCTGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 234
QY 601 TGGGACGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 660
Db 235 TGGGACGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 294
QY 661 CTGCGGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 720
Db 295 ATCAACGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 354
QY 721 GGGTTGCGGCGGCGGTCTGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 780
Db 355 GGGTTGCGGCGGCGGTCTGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 408
QY 781 TACGAGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 840
Db 409 TCCAGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 468
QY 841 GACCGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 900
Db 469 GACCGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 528
QY 901 GCGGAGTGTGCGGAGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 960
Db 529 GCAGAGTGTGCGGAGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 588
QY 961 GTACGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1020
Db 589 GTACGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 648
QY 1021 CTTGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1080
Db 649 TGAAGTGTGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 708
QY 1081 TTCGAGATCAGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1140
Db 709 TTCGAGATCAGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 768
QY 1141 TAGCGATCGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1200
Db 769 TAGCGATCGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 828
QY 1201 TTCGAGATCAGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1260
Db 829 TTCGAGATCAGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 888
QY 1261 GTACCTGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1320
Db 889 GTACCTGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 948
QY 1321 AACTGGAAGTACGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1380
Db 949 AACTGGAAGTACGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1008
QY 1381 AACTGGAAGTACGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1440
Db 1009 AACTGGAAGTACGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1068
QY 1441 ACCACCGCGGTGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1500

Db 1069 AGCACGCGGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1128
QY 1501 ACCGAGTGTGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1560
Db 1129 ACAGTCAACGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1188
QY 1561 ATGCGTGGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1620
Db 1189 ATTTCATGGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1248
QY 1621 CGGACACGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1680
Db 1249 CGGACATACGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1308
QY 1681 CTGCGGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1740
Db 1309 CAAGCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1368
QY 1741 CGATCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1800
Db 1369 AGATCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1428
QY 1801 GGGACACGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1860
Db 1429 GGGACACGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1488
QY 1861 TACCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1920
Db 1489 TATCGGAGCATGCGCGCGGTCTCCAGGAGCGCGCGGAGCATCGCC 1548
QY 1921 ATC 1923
Db 1549 ATC 1551

RESULT 5

ABL54834
ID ABL54834 standard; DNA; 1467 BP.XX AC ABL54834;
XX AC

DT 01-JUL-2002 (first entry)

XX DE Levan fructotransferase encoding sequence.

XX XX Levan fructotransferase; LfTase; difructose dianhydride; DFA IV;
KW levan; gene; ds.

XX OS Arthrobacter nicotinovorans.

XX FH Key Location/Qualifiers
FT CDS 4..1461
FT /*tag= a
FT /product= "levan fructotransferase"

XX JP2002017366-A.

PN 22-JAN-2002.

XX PF 06-JUL-2000; 2000JP-0205756.

XX PR 06-JUL-2000; 2000JP-0205756.

XX PA (NIP) NIPPON TENSAN SEITO KK.
PA (KAGA-) KAGAKU GIUTSU SHINKO JIGYODAN.

XX DR WPI; 2002-287313/33.

XX DR P-PSDB; ABB10073.

XX PT A new levan fructotransferase

XX PS Claim 3; Fig 2; 15pp; Japanese.

XX

CC The present sequence encodes *Arthrobacter nicotinovorans* levan
CC fructotransferase. The present invention also describes a method
CC for the preparation of levan fructotransferase in which a transformant
CC is cultured in a medium and levan fructotransferase is collected from
CC the culture. The method can prepare levan fructotransferase in a
CC large amount.
XX
SQ Sequence: 1452 BP; 299 A; 491 C; 416 G; 246 T; 0 other;

Query Match: 45.5%; Score 909.6; DB 20; Length 1452;
Best Local Similarity 77.0%; Pred. No. 1.4e-136;
Matches 1123; Conservative 0; Mismatches 329; Indels 6; Gaps 1;

QY	466	CGGGCTCGCTCGTGGCGTCTACACATGAGCGCCCCAGCGGCTGGCTCTGGGACCCC	525
DB	1	CAGGCATCCCTCGGGGATCTACACATGACCGCGCCCTCGGCTGGCTATGTGATCGG	60
QY	526	CAACGCCGCTGACACCCAGCGGCTACAGCTGTACTGTACTGTCCGACCGACCAAC	585
DB	61	CAGGACCCCTACATAAAGCGGCTACAGCTGTACTGTACTGTCCGACCGACCAAC	120
QY	586	AACGGCCCCGCGCTGGGACCGAGGACGACGCGGCTGCGCTTTCAGCGCACAC	645
DB	121	AACGGACCGGGGATGGGACCGAGGACGACGCGGAGTGTCTTACACCCACCAT	180
QY	646	GGACCGTGATGGCTGCGCGCGGACTTCCCGTGTGGTCCGGGTGCTGCGGC	705
DB	181	GGAGTGGTGATGCAATGCAACCGGAGTTCCTCGGTGTGGGATCGGCGAGTAGTGGAC	240
QY	706	ACCGGAACACGCGGTTGCGCGCGGCTGCTGCGCTGCGGACCGACCGACCGAC	765
DB	241	ACGCGCAACACCGCGCTTGGCGCGCGGCTGCTGCGCTGCGGACCGACCGACCGAC	300
QY	766	GACGGGCTCCGAAAGTACAGGAGCTACTCTACTGTGTCGACCGAGCGGCTTCAACG	825
DB	301	GAGC-----GAAATTCAGGAACAGTACTTTACTGTGTCACCGGATGCGGCTACTCC	354
QY	826	TTACCGCCCTGCGCGACCGCTATCGTCAACACGACGGTGGCGCGGACCGACCGCC	885
DB	355	TTACCGCGATTTGCTGACCGCGCTATTTGAACACTGATGGAGCGGACCGACCGCCC	414
QY	886	GCGGAGATCAGAACCGCGAGTGGTTCGCGGACCGCCAGATCCACTGGGACCGCGCCG	945
DB	415	GCGGAGTGGAGAACCGAGATGGTTCGCGGACCGCCGAAATTCATGGGAGCGACCGCG	474
QY	946	GGAGATGGGTCTGCGTATCGGACGACTGCGGTACGCGGCTTCTACACTCGCGGAAC	1005
DB	475	AACGAGTGGGTCTGTGTCATCGCAGGCGCGCTACGCTGCTTCTACACTCTCCCAAC	534
QY	1006	CTGCGGACTGGACACTTCCCGCACTTGGACTACCGAACCGCCCTCGGCGGATC	1065
DB	535	CTGCGGATTTGGCAATGGAATGCAATTCGACTACCGCAACCGCCCTCGGCGGTATC	594
QY	1066	GAGTGGCCCGACCTGTTCGAGATCAGCGAGACGGGACCGGACCGGCTGGTCTGCC	1125
DB	595	GAATTCGGGATCTGTTCGAAATGACCGAGGAGAGGAAACCGCGGCTGGGTTCGGG	654
QY	1126	GCCAGCATGACGCGCTACGCGCTCGGCTCCCATGACGTACCGCTTCTACAGCAGCACC	1185
DB	655	GCGAGCATGACGCGCTTACAGCATCGGCTTGGCCATGACCTTTCCCTTACTGACAGTTCA	714
QY	1186	TGGGAGCGGAGAGTTCCACCGGAGGACCTCACCGCGAATGGCTGCACTGGGCTGG	1245
DB	715	TGGAAGGACACACATTCATTCGCGGCAACCTCACACACAGTGGCTTACTGGGATGG	774
QY	1246	GACTGTGACGCGGCTGCTACCTGCGCATGATCGACGCGCGCGGACCGACCGCTCGCC	1305
DB	775	GACTGTGACGCGGCTGCTGACCTGGCGGCTGGAGAGACCTGAGACCAAGCGGCTTGGC	834
QY	1306	ATCGCGTGGATGAACACTGGAAATACGCGGACGCGGAGCTTCCCGACCGACGATCCGAC	1365
DB	835	ACAGCGTGGATGAACACTGGAAATATGCGCGCGCGCAACGCTGCGCCACGCGGCTCCGAT	894

QY	1366	GGCTAACACGGGAGAACTCGATCGTCCGGAGCTCGGGCTCGCCCGACAGCCTGGCGGC	1425
DB	895	GGCTATAACGGGCAAAATTCATCACGCGGAGCTCAGGCTCGAGCCCAATTCGGCGGC	954
QY	1426	TGCTACACCTCTCTGAGCACCCCGTGGCAGCGCTGACGAACCTACGTCACCGCACCAAC	1485
DB	955	TGCTACACCTTGTCTACGACGCGGTTCCCGGCGTTTCGAACTATGCACTCCAGCACC	1014
QY	1486	ACACTCCCGACCGGACCGTTCGAGCGGAGCGGCTCTCTGCGCATGGAACGACGCGCATAC	1545
DB	1015	ACCTTCGCGACCGCACAGTCAACGGCAGTTTCGTACTTCCGTGGAGCGCGCGCGCTAT	1074
QY	1546	GAGATCGAGCTCGACATCGCTGGGACACCGGACGAGCTGCGGCTCGGCTGGCGGC	1605
DB	1075	GAACTGGAACCTGATATTTTCATGGGACACGCGACGCAACCTGCGGTGGCGGC	1134
QY	1606	TCCCGGACGGAACCGGACGACGACATCGGCAAGTACGAGACGAGCTGTAGTCTGAC	1665
DB	1135	TGCTCCGATGGACCGCCCATACGACATCGGCNAATACGCTGACGAGTTGTAGTCTGAT	1194
QY	1666	CGAGGACCTTCGACCTCGCGGGTACTCGCTCGCCCGCTTACTCGGAGCGCGCGCC	1725
DB	1195	CGCGCATCTTCGGAGCAAGCGTTATGCTGCGACCTTACACCGCGCGCGCGCC	1254
QY	1726	ATGACCGCGCGCGCGATCCGTCGACCTGCGGATCTCTGTCGACACCGAGAGCTGAG	1785
DB	1255	ATCGATCGCAACCGAGATCCGTCACCTCGCATCTTTGTAGACACCCCAAGTGTGAG	1314
QY	1786	GTCTTCGTCAACCGCGCCACACGCTCTCTCCAGCAGGTCCACTTCGCGGAGGCGAC	1845
DB	1315	GTGTTGTAATTCGCGGACACGCTGTTTCGACGAGCTGCACTTCGCGCGCGGAG	1374
QY	1846	ACGGGAATCTCGCTTACACCGCGCGCGCGCGCATCTTACCGGCTCGTCTCGTCCG	1905
DB	1375	ACGGGATCTCTCTATGCGGACGCGGCTCGCGGCACTTCACCGGATCACCATCCG	1434
QY	1906	GAGATTGGCGAGGATC	1923
DB	1435	GAGTTGGGAAACCCCATC	1452

RESULT 7
AAD17184
ID AAD17184 standard; DNA; 65140 BP.
XX
AC AAD17184;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nysI DNA of nystatin PKS gene cluster.
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; nysI; ds.
XX Streptomyces noursei.
XX
FH Key Location/Qualifiers
CDS complement (1..1035)
FT /*tag= a
FT /product= "NysD2 partial protein"
FT /note= "CDS does not include stop codon"
FT /*tag= b
FT /product= "NysD1 protein"
FT 2806..6906
FT /*tag= c
FT /product= "NysA protein"
FT 6952..16530
FT /*tag= d
FT /product= "NysB protein"
FT 16550..49840
FT /*tag= e
FT /product= "NysC protein"

```

FT CDS 50260..51015
FT /*tag= f
FT /product= "NysE protein"
FT CDS 51405..54305
FT /*tag= g
FT /product= "NysR1 protein"
FT CDS 54329..57190
FT /*tag= h
FT /product= "NysR2 protein"
FT /*note= "CDS does not include start codon"
FT CDS 57180..59963
FT /*tag= i
FT /product= "NysR3 protein"
FT CDS 60415..61047
FT /*tag= j
FT /product= "NysR4 (short) protein"
FT /*note= "CDS does not include start codon"
FT CDS 61736..62497
FT /*tag= k
FT /product= "NysR5 protein"
FT /*note= "CDS does not include start codon"
FT CDS 63765..64961
FT /*tag= m
FT /product= "ORE1 protein"
XX
XX WO200159126-A2.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
XX
XX 10-APR-2000; 2000GB-0008786.
XX
XX 14-APR-2000; 2000GB-0009387.
XX
XX (UYNØ-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX
XX (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
XX
XX (ALPH-) ALPHARMA AS.
XX
XX (SINV-) SINVENT AS.
XX
XX (DZIE/) DZIEGLEWSKA H.
XX
XX (ZOTC/) ZOTCHEV S B.
XX
XX (SEKU/) SEKUROVA O N.
XX
XX (FJAE/) FJAEVIK E.
XX
XX (BRAU/) BRAUTASET T.
XX
XX (STRO/) STROM A R.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
XX WPI; 2001-557614/62.
XX
XX P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
XX AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
XX useful as antibiotics and antifungals -
XX
XX Claim 2; Page 116-151; 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrolide antibiotic nystatin.
XX The nystatin PKS is useful as antifungal antibiotics. The present
XX sequence is a Streptomyces noursei nysI DNA of nystatin PKS gene cluster.
XX
XX Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;
XX
XX Query Match 7.4%; Score 147.6; DB 22; Length 65140;
XX Best Local Similarity 45.4%; Pred. No. 2.3e-15;
XX Matches 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;

```

```

QY 8 ACCCGGACTTCCCTCGACGACACCGTCCCTCCCTACCGGCGGACCGCGCGGCGGACTGC 67
  || || || || || || || || || || || || || || || || || || || || ||
Db 19803 ACGGCGCGACACCGACCCCGCGCTTCGAGTGCACCCGCGCTGCTGAGCGCGCAC 19862
QY 68 TCTTAGCCTAGACGGGCCCCCTCTCGAGGTCTTCGTCGGGACGAGTGAGGCGACTGCGT 127
  || || || || || || || || || || || || || || || || || || || || ||
Db 19863 AACACGCCCGCCCTACGCCGACCTCGGCGCCATCAGCGCGGCGGCTTCCGCT 19922
QY 128 CGAACCTGGTCTCTCTGGGGCGCGGTGTGACCGCGAGCCTCGAGACGGACGCGCCAG 187
  || || || || || || || || || || || || || || || || || || || || ||
Db 19923 GGAAGGCGTCTGCTCGCGCGCGCGCGCCACACCGTCCGCGCCCGATCGCCCGG 19982
QY 188 GAACCGTGACGTGACCGCGATCGAGTCTGAGCGCGCCAGCGATGCTGACCCCTGAAC 247
  || || || || || || || || || || || || || || || || || || || || ||
Db 19983 CGGCGGACGACCGCTGACCATCGCGTCTACGACCGCGCGCGACCGTCTGCTCGG 20042
QY 248 CTGCGCGCGTCTTGGGCTGACGAGCGCTCCACCGCGGACGAGTCTCTCTTACCGTGC 307
  || || || || || || || || || || || || || || || || || || || || ||
Db 20043 TCGACTCCCTGCTCTCCCGGAGGTCCCGCGCGACGACCCGCGCGCGCGCGCTCC 20102
QY 308 CGAACCCAGGTGGACGCTTCTGTCGCGCGCCACCGCTCCACGAGAGGAACAGCAATGACG 367
  || || || || || || || || || || || || || || || || || || || || ||
Db 20103 ACGGACTCCCTCTTCACGTCGAGTGACCCCGCTCCAGGCGCGCGCGCGCGCGCAC 20162
QY 368 CGGCCATCTACCGCGCGCGCTGCTCCAGGAGCGCGCGCGCGCGAGACTCGCCCTGATCT 427
  || || || || || || || || || || || || || || || || || || || || ||
Db 20163 CGGCCA-----CCGTCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCC 20213
QY 428 TCGGCGGTGCTGTGCGGCTGACGCGCGCGGATCGCTCCGCGGCTGCTCGGCGGTCT 487
  || || || || || || || || || || || || || || || || || || || || ||
Db 20214 TCCGCG-----CCACCGCATCGGACCGCGCGCGCGCGCGCGCGCGCGCGCTCG 20264
QY 488 ACCACATGACCGCGCGCGCGCTGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 547
  || || || || || || || || || || || || || || || || || || || || ||
Db 20265 CGACCGCGAAGGCGCGCTCCCGGACTGCTGTCACACCGCTCAC---CACACCGCGG 20321
QY 548 GCGCCTTACAGCTGTACTACTCTCGACCTCGGACCGAGAACCGCGCGCGCGCGCGCG 607
  || || || || || || || || || || || || || || || || || || || || ||
Db 20322 GCGCGCGCTCCCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20381
QY 608 ACGGAGACGACCGCGCGCTGCGCTTTCAGCAGCAGCGCGCGCGCGCGCGCGCGCG 667
  || || || || || || || || || || || || || || || || || || || || ||
Db 20382 AGTGGCTCCCGCGACCGCGCTTTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 20441
QY 668 CCGACTTCCCGCTGCTGCGGCTCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCG 727
  || || || || || || || || || || || || || || || || || || || || ||
Db 20442 CCGACGGACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20501
QY 728 GCGCGCGCGGCTGCTGCGGCTCGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 787
  || || || || || || || || || || || || || || || || || || || || ||
Db 20502 AGAACCGCGCGGCTTTCGCGCTCTCGACCTCGCGCGCGCGCGCGCGCGCGCGCG 20561
QY 788 ACGAGTACTCTTACTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 847
  || || || || || || || || || || || || || || || || || || || || ||
Db 20562 AGACCTGCGCCACCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20618
QY 848 TCATCGTCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 907
  || || || || || || || || || || || || || || || || || || || || ||
Db 20619 GCACCGAGTGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20678
QY 908 GGTTCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 967
  || || || || || || || || || || || || || || || || || || || || ||
Db 20679 GGAACCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20738
QY 968 GACGACTCGGCTACCGCGCGGTTCTACACCTCCCGCGCGCGCGCGCGCGCGCGCG 1027
  || || || || || || || || || || || || || || || || || || || || ||
Db 20739 CCGCGACCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20798
QY 1028 GCACTTCGACTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1087
  || || || || || || || || || || || || || || || || || || || || ||
Db 20799 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20858

```

```
QY 1088 TCACCGGACGACGAGGACACGCGACCTGCTGCTGCGCGCCAGCATGGACGCTACGGCA 1147
Db 20859 AATCGCGCGCTGCGAGCTGCGGACCGCGCGCTCGCGACCTGCTCGGACGCTCC 20918
QY 1148 TCGGCTCCCATGACGTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT 1207
Db 20919 CGGCGGCGGACCGCGCTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20978
QY 1208 CCGACGACCTACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1267
Db 20979 TCGGCTCCCTACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 21038
QY 1268 GGCATCGATCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1327
Db 21039 GGCACCTGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21098
QY 1328 AGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1387
Db 21099 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21158
QY 1388 TCGTCCGCGAGCTGCGGCTGCG--CGACAGCTGCGGCTGCTGCTGCTGCTGCTGCTG 1445
Db 21159 ACGCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21218
QY 1446 CCCGCTGGACGCGCTGACGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1505
Db 21219 CGTGGACCGGAGCGCTGGCATGAC--AGCAGCGCTGCTGCGACCTGCGACGCTGCG 21276
QY 1506 CGACGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1565
Db 21277 CGCCCGCTCGCGCATGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 21336
QY 1566 CTGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1622
Db 21337 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21396
QY 1623 GCACACGAATCGGCAAGTACGAGCAGACGACCTGCTGCTGCTGCTGCTGCTGCTGCT 1682
Db 21397 CAGGCGGACATCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 21456
QY 1683 CGCGCGGTACTGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1736
Db 21457 CGCGCGAGTCTGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21516
QY 1737 CGCGCGATCGTGCATCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1796
Db 21517 CGCGCGCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 21576
QY 1797 CGCGCGCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1856
Db 21577 CGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21636
QY 1857 GCTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1904
Db 21637 CACCGCGCTGCACTCGGCAACCGCTGACACCGCGCGCGCGCGCGCGCGCGCGCG 21684
```

RESULT 8

AD17186
ID AAD17186 standard; DNA; 125401 BP;

AC AAD17186;

DT 29-NOV-2001 (first entry)

DE Streptomyces noursei nystatin PKS gene cluster DNA.

KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; ds.

OS Streptomyces noursei.

XX key Location/Qualifiers

```
FT CDS 6337..34771  
FT /tag= a  
FT /product= "NysI complete protein"  
FT 34792..51099  
FT /tag= b  
FT /product= "NysJ protein"  
FT 51155..57355  
FT /tag= c  
FT /product= "NysK protein"  
FT 57503..58687  
FT /tag= d  
FT /product= "NysL protein"  
FT complement (58786..58980)  
FT /tag= e  
FT /product= "NysM protein"  
FT /note= "CDS does not include start codon"  
FT /tag= f  
FT /product= "NysN protein"  
FT /note= "CDS does not include start codon"  
FT complement (60238..61296)  
FT /tag= g  
FT /product= "NysD2 complete protein"  
FT 120628..121308  
FT /tag= h  
FT /product= "NysR4 (long) protein"  
XX  
PN W0200159126-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 08-FEB-2001; 2001WO-GB00509.  
XX  
PR 08-FEB-2000; 2000GB-0002840.  
PR 10-APR-2000; 2000GB-0008786.  
PR 14-APR-2000; 2000GB-0009387.  
XX  
XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.  
PA (SNTF-) SINTEF STIFTELSEN IND TEK FORSK.  
PA (ALPH-) ALPHARMA AS.  
PA (SINV-) SINVENT AS.  
PA (DZIE-) DZIEGLEWSKA H.  
PA (ZOTC-) ZOTCHEV S B.  
PA (SEKU-) SEKUROVA O N.  
PA (FJAE-) FJAEVRIK E.  
PA (BRAU-) BRAUTASET T.  
PA (STRO-) STROM A R.  
XX  
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;  
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;  
XX  
XX WPI: 2001-557614/62.  
DR P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,  
DR AAE10149, AAE10150.  
XX  
XX New nystatin polyketide synthase polynucleotides and polypeptides,  
XX useful as antibiotics and antifungals -  
XX  
XX Claim 1; Page 188-254; 266pp; English.  
PS  
XX  
XX The present invention relates to the cloning and sequencing of the gene  
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme  
XX involved in the biosynthesis of the macrolide antibiotic nystatin.  
XX The nystatin PKS is useful as antifungal antibiotics. The present  
XX sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.  
SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;  
Query Match 7.4%; Score 147.6; DB 22; Length 125401;  
Best Local Similarity 45.4%; Pred. No. 2.2e-15;  
Matches 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;
```

QY 8 ACCCGGACTTCCCTCGGACGACCGTCCCTACCGCGCGCGCGCGCGCGCGCGCGCGCTGC 67

Db 80064 ACGGCGGACACACACCGCGCGCTTCGGACTGCACCGCGCTGTGGACGCGCAC 80123
QY 68 TCCTCAGCTACAGCGGCGCTCTCTCGAGTCTTCGTGGGACGAGTGGAGTGGCT 127
Db 80124 AACAGCGCGCGCTACCGCGACCTTCGGGCGCATAGCGCGCGCGCTTCGCTTCGCT 80183
QY 128 CGAAGCTGGTCTCTCGGGGCGGCGGTGTGACCGGAGCCTCGAGAGCGGACGCGCCAG 187
Db 80184 GGAAGGCGTCTCGCTCGCGCGCGCGCGCCACACCGTCCGCGCGCGGATCGCGCCGG 80243
QY 188 GAACCGTGCAGCTGACCGGGATCGAGCTGAGGCGCGCCAGCATGCTGAGCGCCCTGAC 247
Db 80244 CCGGCGAGGACACCTCACCATTGCGCTGTACGACCGCGCGCGCGACCGTGTGTCCG 80303
QY 248 CTGCGCGCGTCTCTGGCTGACGAGCGCTCCACCGCGACAGCTCTCTTACCGCTGCC 307
Db 80304 TCGACTCCTGTCTCTCGCGAGTCTCCCGCGAGTCCCGCGAGCACCAGCGCGCGCGCGCTCC 80363
QY 308 CGAAGCGGTGGAGCTTCGTGCGCGCGACCGCTCCAGAGAGGAACAGCAATGACGC 367
Db 80364 ACCGAGCTCTCTTCCAGCTGAGTGGACCGCTCCAGGCGCGCGCGCGCGCGCGCAC 80423
QY 368 CGGCGATCTACGCGCGCGCTGTCTCAGGAGAGCGCGCGCGAGCACTCGCGCTGTATCT 427
Db 80424 CGGCGA-----CCGTGCGCGTCTCTGCGCGCGAGCGCGCGCGCGCGCGCGCGCAC 80474
QY 428 TCGCGCGTGTGTGCGCGCTGACGCGCGGCGATCGCTCCGCGGCTCGCTCCGCTGCTCT 487
Db 80475 TCGCGG-----CCAGCGCATCGGAGCACCAGCGCGCGCGCGCGCGCGCGCGCTCG 80525
QY 488 ACCATGACGCGCGCGCGCTGGCTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 547
Db 80526 CCGAGCGCGAGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80582
QY 548 GCGCTTACAGTGTACTACCTGCTGCTCGGACGAGCAACAGCGCGCGCGCGCGCGCGCG 607
Db 80583 GCGCGCGCGTTCG 80642
QY 608 ACGGAGCAGCAGCGCGCGTTCGCTTCACGACACAGCGCGCGCTGATGCGCGTTCGCG 667
Db 80643 AGTGGCTCGCGCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80702
QY 668 CGGACTTCCCGTGTGCTCGGCGTGGGCGTGTGCGGACGCGCGCGCGCGCGCGCGCGCG 727
Db 80703 CGGAGCGCACCG 80762
QY 728 GCGCGCGCGGTGCTGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 787
Db 80763 AGAAGCGCGCGGTTCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80822
QY 788 AGCAGTACTCTACTGCTGCGCGCGCGCGGTTCACGTTACCGCGCTGCGCGCGCGCGCG 847
Db 80823 AGACCTTGGCCACCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80879
QY 848 TCATCGTCAACACCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGT 907
Db 80880 GCACCGAGTGCACCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80939
QY 908 GGTTCGCGGACCGCAAGATCTACTGGGACACCGCGCGCGCGCGCGCGCGCGCGCGCG 967
Db 80940 GGAACCGCGGAGCGCGCGCTGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80999
QY 968 GAGGACTCGGTACCGCGCTTACACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1027
Db 81000 CCGCGCGCTGTGCG 81059
QY 1028 GCACTTCGACTACCGGACCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1087
Db 81060 CGGCG 81119
QY 1088 TCACCGCAGACGCGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1147

Db 81120 ACATCGCGCGCTTCGCGAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81179
QY 1148 TCGGCGTCCCGATGAGCTACGCTACTGACAGGACACTTGGAGCGGAGAGTTCACAG 1207
Db 81180 CGGCGGGGACCGCTCACCCTGCTGCTCCACACCGCGCGCGCTGCTGAGACGAGCGGTCC 81239
QY 1208 CGGAGGACCTACCGCGCAATGGCTGACTGGGGTGGGACTGGTACGCGGCGCGTCACT 1267
Db 81240 TCGGCTCCCTACCGCGAGCGCTTGGACACCGTCTCTGCGCGCGCGCGCGCGCGCT 81299
QY 1268 GGGCATGATCAGCG 1327
Db 81300 GGCACCTGACGAGCG 81359
QY 1328 AGTAGCG 1387
Db 81360 TCG 81419
QY 1388 TGTGTCGCGAGTGGCGCTGCG--CGGACAGCTTGGCGGTGGTACACCTCTCTGAGCAC 1445
Db 81420 ACGCGCTCG 81479
QY 1446 CCCGCTGGCAGCGCTGACGAACTAGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1505
Db 81480 CGTGGACCG 81537
QY 1506 CGACGCGCGCGCGCTGCGCATGGAACGCGCGCATACGAGATCGAGCTCGACATCGC 1565
Db 81538 CGCGCGCTCG 81597
QY 1566 CTGGGACACCG 1622
Db 81598 CCG 81657
QY 1623 GCACGAGACATCGGCAAGTACGAGCAGCAGCTGACGTGCGCGCGCGCGCGCGCGCG 1682
Db 81658 CAGGCG 81717
QY 1683 CGCGCGGTACTCG 1736
Db 81718 CGCGCGGTCTCGAGACCG 81777
QY 1737 CGCGCGGTCTCG 1796
Db 81778 CG 81837
QY 1797 CG 1856
Db 81838 CG 81897
QY 1857 GGTCTACACCG 1904
Db 81898 CACCGCGCTGAGCTCCG 81945

RESULT 9

AA53491/C

ID AAX53491 standard; DNA; 114955 BP.

XX AAX53491;

AC AAX53491;

XX AAX53491;

DT 05-JUL-1999 (first entry)

XX Human adenosine A1 receptor antisense oligonucleotide fragment.

XX Antisense oligonucleotide; multiple target; antisense treatment;

KW impaired respiration; inflammation; lung disease;

KW pulmonary vasoconstriction; inflammation; allergic rhinitis;

KW acute asthma; allergy; asthma; impeded respiration;

KW respiratory distress syndrome; pain; cystic fibrosis;

KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;

KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;

KW colon cancer; breast cancer; lung cancer; pancreatic cancer;

KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW prostate cancer; ss.
OS Synthetic;
XX WO9913886-A1.
PN
XX
PD
XX
PF 25-MAR-1999.
XX 17-SEP-1998; 98WO-US19419.
XX
XX
PR 09-JUN-1998; 98US-0093972.
PR 17-SEP-1997; 97US-0059160.
XX
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
XX
PI Nyce JW;
XX
DR WPI; 1999-229400/19.
XX
XX
PT New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT vasoconstriction
XX
XX
PS Disclosure; Page 37; 120pp; English.
XX
XX
CC The specification describes antisense oligonucleotides (AA52869-X55271)
CC directed against at least 2 mRNAs selected from target genes, coding and
CC non-coding regions of RNAs corresponding to target genes, gene
CC initiation codons, genomic flanking regions, intron-exon borders, the
CC 5'-end, the 3'-end and the juxta-section between coding and non-coding
CC regions and all segments of RNAs encoding proteins associated with one
CC or more diseases, conditions or mixtures. The antisense oligonucleotides
CC may be derived from sequences AA5272-74. These multiple target
CC oligonucleotides (specifically AA5180-271) can be used for the
CC antisense treatment of diseases and conditions. Typical diseases and
CC conditions are those associated with impaired respiration and
CC inflammation, including lung diseases, pulmonary vasoconstriction,
CC inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded
CC respiration, respiratory distress syndrome, pain, cystic fibrosis,
CC pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic
CC obstructive pulmonary disease (COPD), and cancers such as leukemias,
CC lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,
CC pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,
CC hepatic metastases, as well as all types of cancers which may metastasize
CC or have metastasized to the lungs, including breast and prostate cancer.
XX
XX
SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;
Query Match 6.7%; Score 134.8; DB 20; Length 114955;
Best Local Similarity 31.6%; Pred. No. 2.4e-13;
Matches 617; Conservative 180; Mismatches 1134; Indels 20; Gaps 5;
QY 32 CCGTCCCTACCGGCGACCGCCGCGGCGGCTGCTCTCAGCTAGACGGGCGCCCTCC 91
DB 105241 CCGGCGCGCGCCNNHNNNSGCGCGCGCGCGCGCGCGCCNNHNNNSGCGCGCGCG 105182
QY 92 TCGAGTCTTCGTGCGGGACGCTGAGGCGACTCGCTGCAACCTGPTCTCTTGGGGCGG 151
DB 105181 GCGGCGCGCCVNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105124
QY 152 CGGCTGTACCGGCGGCTCGAGACGGACGGCCAG-GAACTGTGACGTGACCGCGATC 210
DB 105123 CCGGCG 105064
QY 211 GAGCTGAGCG 270
DB 105063 CCGGCG 105004
QY 271 GCGCTCCACCGACGACTCTCTCTTCTTACCGCTGCGCGGAAACAGGCTTCTGCTC 330
DB 105003 NNHNNNSGCG 104944
QY 331 GCGCCACCGCTCCACGAGAGGACCAATGACCGCGCGCGCGCGCGCGCGCGCGCG 390

DB 104943 GCCCVGCGCGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104884
QY 391 CTCCAGGAGCGCGCGCGGAGCACTCGCCCTGATCTTCGGCGGTCTGTGCGCGCTGCA 450
DB 104883 CVGCGNNHNNNSVGGCCVCGCGNNHNNNSVGGCCVCGGNNHNNNSVGGCCVGG 104824
QY 451 GCCCGGCGATCGCTCCGGGCTCGCTCCGTGCGTCTACCATGACGGCCCCCAGCGGC 510
DB 104823 NNHNNNSCCVGGCCVCGGNNHNNNSGCCVGGCCVCGGNNHNNNSGCCVGGCCV 104764
QY 511 TGCGCTCGGACCCCAACGCGCGGTACACCCAGCGCGCTTACAGCTGTACTACTG 570
DB 104763 CGNNHNNNSGCG 104704
QY 571 CACTCCGACCAACAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 630
DB 104703 CGCCVGGCCVCGGNNHNNNSGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104644
QY 631 GCTTTCACGACACAG-GCACCCTGATCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 689
DB 104643 GCVGCGGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104584
QY 690 GTCGGCGGTCTCGGACCGCGGAAACGAGAGGTTCGGCGCGCGCGCGCGCGCGCT 749
DB 104583 VCGGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104524
QY 750 CGCGACCCAGCG 809
DB 104523 CVGCGNNHNNNSGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104464
QY 810 CGAGCGCGGTTCACGTTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 869
DB 104463 CVGCGCGCGGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104404
QY 870 CGCGCGCACCG 929
DB 104403 CG 104344
QY 930 CTGGGACACCGCGCGCGGAGATGGTCTCGCTCATCGGACGACTCGGCTACGCCGCT 989
DB 104343 NNSGCCCG 104284
QY 990 C-----TACACTTCGCGCAACTGCGCGACTGAGCACTTCGCGCGCACTTCGAC 1038
DB 104283 CVGCGCGCGGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104224
QY 1039 TACCGAACCGCGCTCGGCGCGCATCGAGTCCCGGCGCTGTCGAGATCAACCGCAG 1098
DB 104223 CGCGCGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104164
QY 1099 GACGGG-----CACGCCACTGGGTCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1153
DB 104163 GAGNNHNNNSGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104104
QY 1154 TCCCATGACGTACGCTTACTGACAGGCACTGAGCGCGCGCGCGCGCGCGCGCG 1213
DB 104103 CCGCGCGCGGNNHNNNSGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104044
QY 1214 ACCTCACCGCAATGGCTCGACTGGGCTGGGACTGTTACGCGCGCGCTCACTGGCCAT 1273
DB 104043 NNHNNNSGCCCG 103984
QY 1274 CGATGACCGCGCGCGGAGCAAGCGCTCGCATCGCTGGGATGACAACTGGAAGTACG 1333
DB 103983 CCG 103924
QY 1334 CGCGAGCGAGCTCCCG 1393
DB 103923 GCG 103864
QY 1394 GCGAGCTGGGCTCG 1453

Db 103863 CGGCCGACGCGCCCAVAGAGGNNHNNNSGCCGCGCCCGCCCAVAGAGCGGNNHNN 103804
 QY 1454 CAGCGCTGACGAATACGTACCGCCACACACACACTCCCGCAGCGGACCGTCGACGGCA 1513
 Db 103803 NSGCCGCGCCGACGCGCCCAVAGAGGNNHNNNSGCCGCGCCCGCCCAVAGAGC 103744
 QY 1514 GCGCGCTCTGCAATGAGAGCGGCGGATAGAGATCGAGCTCGACATCGCTGGGACA 1573
 Db 103743 GCGGNNHNNNSGCCGCGCCCGCCCAVAGAGGNNHNNNSGCCGCGCCCGCCGAGC 103684
 QY 1574 CCGGACGACGCTGCGCATCTCGTGGCGCGCTCCCGCCGAGCGAACCAGCGCACACGA 1633
 Db 103683 CGCCCAVAGAGCGCGGNNHNNNSAGCGCGCGGNNHNNNSAGCGCGGNNHNNNSV 103624
 QY 1634 TCGGCAAGTACGAGACGACCTGTAGTGTGACGAGGACCTCCCGACCTCGCGGGTACT 1693
 Db 103623 GACGCGCGGNNHNNNSAVAGCGCGCGGNNHNNNSAVAGCGCGCGGNNHNNNSC 103564
 QY 1694 CGTCCGCGCCCTACTCGCGAGCGCGCGCCCGCCCATCGACCGCGCGCGCGATCG 1753
 Db 103563 CAVGAGCGCGCGGNNHNNNSGCAVAGCGCGCGGNNHNNNSGCGCAVAGAGCGCG 103504
 QY 1754 TCGCATCTCTGTCGACACGAGAGCGTCTGCTCAAGCGCGCCACACCGTGC 1813
 Db 103503 CNHNNNSCGCCCAVAGAGCGCGGNNHNNNSGCCGCGCCCAVAGAGCGCGGNNH 103444
 QY 1814 TCTCCAGCAGGTCTCCTTTCGCGAGGCGGACACGCGGATCTCTCTACACCGAGCGG 1873
 Db 103443 GCGGCAVAGCGCGCGGNNHNNNSAGCGCGCAVAGAGCGCGCGGNNHNNNSGAG 103384
 QY 1874 GCGCGCACACTTCACCGGCGATCGTCTCGGAGAGATTCGCGAGGCGATCTAGGCGAT 1933
 Db 103383 GCGCAVAGCGCGCGGNNHNNNSGAGCGCGCAVAGAGCGCGCGGNNHNNNSCGAG 103324
 QY 1934 ACACACACGCTCACGGAAGCGCGCGCGCG 1964
 Db 103323 CGCCCAVAGCGCGCGGNNHNNNSCGCGAGC 103293

RESULT 10
 ID AAF25795
 XX AAF25795 standard; DNA; 3849 BP.
 AC AAF25795;
 XX
 DT 09-APR-2001 (first entry)
 XX
 DE S. chrysomallus acm-C DNA.
 XX
 KW Peptide synthetase; PPS; actinomycin synthetase; acm-C; penicillin;
 KW N-methyltransferase domain; activation domain; amino acid synthesis;
 KW vancomycin; cephalosporin; pristinamycin; actinomycin D; ds.
 XX
 OS Streptomyces chrysomallus.
 XX
 PN DE19928313-A1.
 XX
 PD 21-DEC-2000.
 XX
 PF 16-JUN-1999; 99DE-1028313.
 XX
 PR 16-JUN-1999; 99DE-1028313.
 XX
 PA (KELLER) KELLER U.
 XX
 PI Schauwecker F, Keller U;
 XX
 DR WPI; 2001-081744/10.
 XX
 PT Introducing N-methyltransferase activity into peptide synthetase
 PT enzymes, useful for synthesis of N-methylated peptides such as
 PT antibiotics, by altering DNA -
 XX

PS Example 2; Page 9; 18pp; German.
 XX This invention describes a novel method of introducing an
 N-methyltransferase domain (A) into peptide synthetase (PPS) activation
 domains by altering the DNA that encodes the domains. The invention also
 describes a method for combining genes (or gene segments) that encode PPS
 modules that lack (A) with genes (or segments) encoding modules that
 include (A). Modified DNA sequences formed by insertion of (A) are used
 (i) for altering natural (or already altered) PPS or polyketide
 synthetase (PKS) genes and their fragments, also for construction of new
 PPS and PKS gene and (ii) for construction of plasmids or genetically
 altered organisms for synthesis of encoded proteins (B). (B) are used for
 in vivo or in vitro enzymatic synthesis of amino acids, polypeptides, and
 peptidyl-acetyl mixed structures containing N-methylated amino acids, or
 their derivatives, also for fermentative production of such compounds.
 These compounds are often of pharmaceutical value, e.g. penicillins,
 vancomycin, cephalosporin, pristinamycin or actinomycin D. The modified
 PPS (enzymes involved in non-ribosomal peptide synthesis) are able to
 N-methylate their substrates but substrate specificity remains unchanged.
 XX
 SQ Sequence 3849 BP; 596 A; 1685 C; 1122 G; 446 T; 0 other;
 Query Match 6.0%; Score 119.8; DB 22; Length 3849;
 Best Local Similarity 44.6%; Pred. No. 7.9e-11;
 Matches 558; Conservative 0; Mismatches 687; Indels 6; Gaps 2;
 QY 583 AACACGCGCGCGCGCTGGGACACACGAGCAGCAGCGGCTCGCTTCACGCAC 642
 Db 192 AGCGAGCGCGCGCTAGCTACTCCGAGCTCAACACGCGCGCCACCTCGCCAC 251
 QY 643 CAGGCGACCGTATCGCGCTGGCGCGGCTTCCCGCTGTGTCCGGGTGCGGCTGCTC 702
 Db 252 CAGCTCACCACCGCGCGGCTTCCCGCGGCGAGCGCGCTCTCTTCCAACTCC 311
 QY 703 GGCACCGCGGACACGCGAGGTTCCGCGCGCGGCTGCTCGCTCGCGACCGCGG 762
 Db 312 CCGGACCGCTCACCACCGCTCTCGCCCTCGCCCAAGCGCGCGGCTACATCTCC 371
 QY 763 ACCGCGCGCTCGCGAAGTACCGAGGAGTACTCTACTGTGTCGACGCGCGGCTC 822
 Db 372 GACAGCGCTACCGCGCGCGCTACCGCTCTCTCGAGGAGACCGCGCAACATC 431
 QY 823 AGTTTACCGCGCG ---TGCGCGACCGCTGTCTGTCAACACGCGGTCGCGCGCGCAC 879
 Db 432 CTCATCACCAGCACACACCGCTCGACACCGCGGCTTCAACCGCGCGCAC 491
 QY 880 AGCGCGCGGAGATCGAGACGCGGAGTGTCCCGCGCGCGGAGTCCACTGGGACAC 939
 Db 492 ACCCGCGCGGAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 551
 QY 940 GCGCGCGGAGATGGGTCTGTGCTGTCTGCGAGCTCGGTAGCGCGCTTCTACACTCG 999
 Db 552 GCCTACATGTACACAGCGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCG 611
 QY 1000 CCGAGCTCGGCGAGTGGGACACTTCCCGCGACTTTCGCTACCGCGCGCGCGCGCG 1059
 Db 612 CGCACATCAGCGCGCTCGCGCTCGACCGCGCGCTTTCGCGCGCGCGCGCGCG 671
 QY 1060 GGCATCGAGTCCCGCGCGCTTTCGAGATCAGCGAGAGCGGCGCGCGCGCGCG 1119
 Db 672 CTCCTCCACTCCCGCGCGCTTTCGAGGCTTCCACTAGAGATCTGGGTCCCGCTC 731
 QY 1120 CTCGCGCGCGAGTGGAGCGCTTACGCGCTCGCGCTCCCGCTACGCTACTGGACA 1179
 Db 732 AACGGCAACACCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 791
 QY 1180 GGCACCTGGGAGCGAGTTCACGCGCGAGCTTCCCGCGCGCGCGCGCGCGCGCG 1239
 Db 792 GTCATCAGCGAGCAGATCAGCGCGCTTGTGCTGACGAGTGGGTCTTCAACCTCC 851
 QY 1240 GCGTGGGACTGTACGCGCGCGCTACCTGGCGCTCGATCGACGCGCGCGCGCGCG 1299
 Db 852 ACCGAGCAGCGCGCGGAGCTTTCACCGCGGTCGCGGAGATCTGGAGCGCGCGCG 911

Db 21093 CTGCGCGGCGCGCGCGCGCTACCGAGAGTCTCTGCGCACCATCCGCGCGCGCGG 21152
QY 1689 GTACTCGCTCGT 1748
Db 21153 CGAGGAACCTCTCG 21212
QY 1749 GCACCTCG 1808
Db 21213 GCGCGCTGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21272
QY 1809 CGTGTCTCTCG 1868
Db 21273 CGCGTACACATCCCGCTGACATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21332
QY 1869 CG 1880
Db 21333 CG 21344

RESULT 12

AADI7185

ID AADI7185, standard; DNA: 27541 BP.

XX

AC AAD17185;

XX

DT 29-NOV-2001 (first entry)

XX

DE Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.

XX

KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;

XX

KW antifungal; antibiotic; nys2; ds.

XX

OS Streptomyces noursei.

XX

FH Key Location/Qualifiers

FT CDS complement (454..1191)

FT

FT /*tag= a

FT /product= "NysF protein"

FT /note= "CDS does not include start codon"

FT complement (1275..3092)

FT /*tag= b

FT /product= "NysG protein"

FT complement (3070..4824)

FT /*tag= c

FT /note= "CDS does not include start codon"

FT /product= "NysH protein"

FT 5122..6156

FT /*tag= d

FT /product= "NysD3 protein"

FT 6338..27541

FT /*tag= e

FT /product= "NysI partial protein"

FT /note= "CDS does not include stop codon"

XX

PN W0200159126-A2.

XX

PD 16-AUG-2001.

XX

PF 08-FEB-2001; 2001WO-CB00509.

XX

PR 08-FEB-2000; 2000GB-0002840.

PR

PR 10-APR-2000; 2000GB-0008786.

PR

PR 14-APR-2000; 2000GB-0009387.

XX

PA (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.

PA (SNPF) SINTEF STIFTELSEN IND TEK FORSK.

PA (ALPH-) ALPHARMA AS.

PA (SINV-) SINVENT AS.

PA (DZIE-) DZIEGLEWSKA H.

PA (ZOTC-) ZOTCHEV S B.

PA (SEKU-) SEKUROVA O N.

PA (FJAE-) FJAERVIK E.

PA (BRAU-) BRAUTASET T.

(STRO/) STROM A R.

PA

XX

PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;

PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;

XX

XX WPI: 2001-557614/62.

DR

DR P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.

XX

XX New nystatin polyketide synthase polynucleotides and polypeptides,

PT useful as antibiotics and antifungals.

XX

XX Claim 2: Page 151-166; 266pp; English.

XX

CC The present invention relates to the cloning and sequencing of the gene

CC cluster encoding a modular type I polyketide synthase (PKS) enzyme

CC involved in the biosynthesis of the macrolide antibiotic nystatin.

CC The nystatin PKS is useful as antifungal antibiotics. The present

CC sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.

XX

XX

SQ Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

Query Match

Best Local Similarity 5.5%; Score 109.2; DB 22; Length 27541;

Matches 859; Conservative 43.5%; Pred. No. 3.2e-09;

Mismatches 1088; Indels 27; Gaps 7;

QY 4 GTGACCCCGACTTCCCTTCGAGACACCGTCCCTACCGGCGCGACCGCGCGCGCGA 63

Db

8285 GTCCGCGCGCGACACCTCCGCGCGCGACCTCCGCGCGGAGATCGTCGCGCGCGCGCG 8344

QY 64 CTGCTCTCAGCTAGACGGCGCGCGCTCTCGAGGTCTTCTCGGCGCGCGCGCGACT 123

Db

8345 GGGGTCTCTCCCTGCG 8404

QY 124 GCGTCGAACCTGCTCTCTGGGGCGCGCGGTGTACCGCGCGCGCGCGCGCGCGCG 183

Db

8405 GCGTCCCGCGACGGCGCGCGGCGATGATCGCGTCCAGCGCGTCCGAAACCGCGCGCG 8464

QY 184 CCAGGAACCGTGCAGCTGACCGCGGATCGAGCTCGAGGCGCGCGCGCGCGCGCGCT 243

Db

8465 CTGCTCGCT 8524

QY 244 GAACCTGCGCGCGCTTCTGGGCTGACGAGCGCTCCACCGCGCGCGCGCGCGCGCG 303

Db

8525 GTCTGTCCGGCG 8584

QY 304 TGCCCGAACCGGTGGAGCGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 363

Db

8585 CGCAAGACCG 8644

QY 364 AGCGCGCGCATCTCAGCT 423

Db

8645 CTGCGAGCGCTTCCGCT 8704

QY 424 ATCTTTCGCGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 483

Db

8705 GTCTCCAACTGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8764

QY 484 GTCTACCACTGACCG 543

Db

8765 GTGCGACCGTCCG 8824

QY 544 CACGGCGCGCTACCGAGTGTGTA---CTACTGTGACTCGGACCGCGCGCGCGCGCG 600

Db

8825 GACACCGCGCGCTTCTCGAACTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8884

QY 601 TGGGACCG 660

Db

8885 TGCGTGGAGCG 8944

QY 661 CTGCG 714

Db

8945 GAGGAGCGACCGCTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9004

```

Query Match      5.0%; Score 101; DB 22; Length 4403765;
Best Local Similarity 42.3%; Pred. No. 4.1e-08;
Matches 833; Conservative 0; Mismatches 1115; Indels 23; Gaps 4;

QY 11 CCGACTTCCTCGAGACACCGTCCCTATCCGCCGACCCCGGCCGACTGCTCC 70
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3941672 CGCGCGCGCCCGCGACACCGCGTTCCTGTGTTGCCATATCGCCGGTGCCTCG 3941613

QY 71 TCAGCCTAGACGGGCCCTCCTCGAGGTCVTCGTCGGGGACGGTGAGGGGACTGCGCTCGA 130

```


KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW prostate cancer; ss.
XX Synthetic.
XX WO9913886-A1.
XX PD 25-MAR-1999.
XX PF 17-SEP-1998; 98WO-US19419.
XX PR 09-JUN-1998; 98US-0093972.
XX PR 17-SEP-1997; 97US-0059160.
XX PA (UYEC-) UNIV EAST CAROLINA.
XX PI Nyce JW;
XX DR WPI; 1999-229400/19.
XX PT New antisense oligonucleotides used in treatment of, e.g. pulmonary
XX PT vasoconstriction
XX PS Disclosure; Page 37; 120pp; English.
XX CC The specification describes antisense oligonucleotides (AA52869-X55271)
CC directed against at least 2 mRNAs selected from target genes, coding and
CC non-coding regions of RNAs corresponding to target genes, gene
CC initiation codons, genomic flanking regions, intron-exon borders, the
CC 5'-end, the 3'-end and the juxta-section between coding and non-coding
CC regions and all segments of RNAs encoding proteins associated with one
CC or more diseases, conditions or mixtures. The antisense oligonucleotides
CC may be derived from sequences AA5272-74. These multiple target
CC oligonucleotides (specifically AA55180-271) can be used for the
CC antisense treatment of diseases and conditions. Typical diseases and
CC conditions are those associated with impaired respiration and
CC inflammation, including lung diseases, pulmonary vasoconstriction,
CC inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded
CC respiration, respiratory distress syndrome, pain, cystic fibrosis,
CC pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic
CC obstructive pulmonary disease (COPD), and cancers such as leukemias,
CC lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,
CC pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,
CC hepatic metastases, as well as all types of cancers which may metastasize
XX or have metastasized to the lungs, including breast and prostate cancer.
XX SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;
Query Match 5.0%; Score 99.8; DB 20; Length 114955;
Best Local Similarity 34.0%; Pred. No. 8.9e-08;
Matches 604; Conservative 118; Mismatches 1032; Indels 21; Gaps 7;
QY 2 CGGTGACACCCGACTTCCCTCGACGACACACCTCCCTTACCGCGCGACCGCGCGCC 61
DB 104356 CBGGGCGCGCGCGCGGGSNNNDNCCBGGCBGGCGCGCGCGCGCGCGGGSNNN 104415
QY 62 GACTGCTCTCAGCTAGACGGCGCCCTCTCTCGAGGTCTTCGTGGGGAGGGTGAGCGGA 121
DB 104416 DNNCCGCBGGCGCGCGCGCGGGSNNNDNCCBGGCBGGCGCGCGCGCGCG 104475
QY 122 CTCGTGCAACCTGCTCTCTGGGCGCGCGCGGTGTGACCGGAGCTCTGACAGCGCAC 181
DB 104476 GCCSNNNDNCCBGGCGCGCGCGCGGGSNNNDNCCBGGCBGGCGCGCGCGCG 104535
QY 182 GGCAGGAAACCGTGACGTGACCGCGATGACGTGACGCGCGCGCGCGCGCGCGCG 241
DB 104536 CGCGGSNNNDNCCBGGCBGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104595
QY 242 CTGAACCTCGCGCTCTGGGTGACGAGCGCTCCACCGCGCGCGCGCGCGCGCGCG 301
DB 104596 CGCGGSNNNDNCCBGGCGCGCGCGCGGGSNNNDNCCBGGCBGGCGCGCGCGCG 104655
QY 302 GCTGCCCGAACAGGTGGACGCTTCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 361

DB 104656 SNNNDNCCBGGCBGGCGCGCGGSNNNDNCCBGGCBGGCGCGCGCGCGCGCGCG 104715
QY 362 TGAGCGCGCGCATCTACGCGCGCGCGCGCTGCTCCAGGAGCGCGCGCGCGCGCGCG 421
DB 104716 CGBGGCBGGCGCGGSNNNDNCCBGGCBGGCGCGCGGSNNNDNCCBGGCGCGCG 104775
QY 422 TGATCTTTCGGCGGTGCTGTGTCGCGCGCTCAGCCCGCGCGCATCGCTCCGGGTTCGCTCG 481
DB 104776 SNNNDNCCBGGCBGGCGGSNNNDNCCBGGCBGGCGCGGSNNNDNCCBGGCGCGCG 104835
QY 482 CCGTCTACCATGACCGCGCGCGCGCGGTGTCTCGGACCGCGCGCGCGCGCGCGCGTACCA 541
DB 104836 SNNNDNCCBGGCBGGCGGSNNNDNCCBGGCBGGCGCGGSNNNDNCCBGGCGCGCG 104895
QY 542 CCCACGCGCGCTACCACTGCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 601
DB 104896 CBGGCBGGCG 104955
QY 602 GGGACACGCGAGCAGCAGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 658
DB 104956 GGGCGGSNNNDNCCBGGCBGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105015
QY 659 CGCTGCGCGCGCGCTTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 718
DB 105016 CG 105075
QY 719 CAGGTTTCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 778
DB 105076 CBGGCG 105133
QY 779 AGTACAGGAGCAGTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 838
DB 105134 NNDNCCBGGCG 105193
QY 839 CCACCGCGCTATCTGCTAACACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 898
DB 105194 NNDNCCBGGCG 105253
QY 899 ACCCGAGTGTTCG 958
DB 105254 DNNCG 105312
QY 959 CGCTATCGGAGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1018
DB 105313 CGCGCGCGCGCGCGGSNNNDNCCBGGCBGGCGCGCGCGCGCGCGCGCGCGCGCGCG 105372
QY 1019 CACTTCCCGCACTTCGACTACCGCAACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1078
DB 105373 CGSNNNDNCCGCGCGCGCGCGCGGSNNNDNCCGCGCGCGCGCGCGCGCGCGCGCGCG 105432
QY 1079 TGTTCAGATCACCGCAGACGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1136
DB 105433 GCGCGSNNNDNCCGCGCGCGCGCGGSNNNDNCCGCGCGCGCGCGCGCGCGCGCGCGCG 105492
QY 1137 CGCTACCGCATCGGCTTCCCATGACGTACGCTTACTGACAGGACGCGCGCGCGCGCGCG 1196
DB 105493 NNDNCCBGGCBGGCTCCGCBBSNNNDNCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105552
QY 1197 GCAGTTCCAGCGCAGCACTTACCGCGCAATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1256
DB 105553 GCGCGCGCGCGCGGSNNNDNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105612
QY 1257 GCGCGTCACTGCGCATCGATCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1316
DB 105613 CGCG-----CGSNNNDNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105667
QY 1317 GAACAACTGGAAGTACCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1376
DB 105668 CCSSNNNDNCCG 105727
QY 1377 GCAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1431

```
Db 105728 GGGCCCCCTGGCTGGCCCCCGCGCGCGCTTGGCCCGCCGCSNNNDNNGGCCCCCTGGC 105787
Qy 1432 ACCCTCTCTGAGCACCCCGTGGCAGCGCTGACGAACCTAGCTACCGCCGACACACACACTC 1491
Db 105788 TCGGCCCCCGCGCGCGCTTGGCCCCCGGCSNNNDNNGGCCCCCTGGCTCGGCCCGCGG 105847
Qy 1492 CCGGACCGGACCGTGGAGCGGAGCGCGCTCTCTGCCATGG---AACGGAGCGGCATACGAG 1548
Db 105848 CCGGCTTGGCCCGCCGCSNNNDNNGGCCCCCTGGCTCGGCCCGCGCGCGCTTGGCCG 105907
Qy 1549 ATCGAGCTGACATCGCTGGGACACCGGAGACGAGTGGCGATCTCGTGGGCGGCTCC 1608
Db 105908 CCCNNNDNNGGCGCCCTGGCTCGCGCCCGCGCGCTTGGCCGCSNNNDNNGGCGC 105967
Qy 1609 CCGGACGGAACCGGACACGACGCAATCGGCAAGTACGAGCAGACCTGCTAGTCGACCGA 1668
Db 105968 CCGTGGCTGGCCCCCGCGCGCGCTTGGCCGCSNNNDNNGGCCCCCTGGCTCGGCCCGCGC 106027
Qy 1669 GGACCCCTCGGACCTCGCGGGGTACTCGCTCGCCCCCTACTCGGAGCGCGCGCCCCCATC 1728
Db 106028 GGCCCGGCTTGGCCGCSNNNDNNGGCCCCCTGGCTCGGCCCGCGCGCGCTTGGCCSNN 106087
Qy 1729 GACCCCGCGCGCGCATCGCTGCACCTGCGCATCCT 1763
Db 106088 NDNNGGCGCCCTGGCTCGGCCCCCGCGCGCGCTT 106122

RESULT 15
AA14651
ID AA14651 standard; DNA; 77536 BP.
AC AA14651;
XX
XX
XX 08-AUG-2000 (first entry)
XX
XX Nucleotide sequence of the FK-520 biosynthetic gene cluster.
FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
Streptomyces hygroscopicus var. ascomyceticus; immunophilin;
FK-506 binding protein; polyketide compound; transplant rejection;
graft-versus-host disease; uveitis; alopecia universalis;
autoimmune chronic active hepatitis; inflammatory bowel disease;
multiple sclerosis; primary biliary cirrhosis; scleroderma;
neutrite outgrowth; nerve regrowth; Parkinson's disease;
Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
peripheral neuropathy; ss.
Streptomyces hygroscopicus.
FH Key | Location/Qualifiers
FT CDS complement (412..1836)
FT /tag= a
FT /note= "fkBW gene"
FT complement (2020..3579)
FT /tag= b
FT /note= "fkBV gene"
FT 3969..4496
FT /tag= c
FT /note= "fkBR2 gene"
FT complement (4595..5488)
FT /tag= d
FT /note= "fkBR1 gene"
FT 5601..6818
FT /tag= e
FT /note= "fkBE gene"
FT 6808..8052
FT /tag= f
FT /note= "fkBF gene"
FT 8156..8824
FT /tag= g
FT /note= "fkBG gene"
FT complement (9122..9883)
FT /tag= h

/note= "fkBH gene"
complement (9894..10994)
/tag= i
/note= "fkBI gene"
complement (10987..11247)
/tag= j
/note= "fkBJ gene"
complement (11244..12092)
/tag= k
/note= "fkBK gene"
complement (12113..13150)
/tag= l
/note= "fkBL gene"
complement (13212..23988)
/tag= m
/note= "fkBC gene"
complement (13452..13662)
/tag= n
/note= "ACP6"
complement (13761..14394)
/tag= o
/note= "KR6"
complement (14517..15294)
/tag= p
/note= "ER6"
complement (15438..16587)
/tag= q
/note= "dehydratase domain (DH) 6"
/tag= r
complement (16587..17820)
/note= "acyltransferase domain (AT) 6"
complement (17820..19053)
/tag= s
/note= "KS6"
complement (19116..19326)
/tag= t
/note= "ACP5"
complement (19464..20097)
/tag= u
/note= "KR5"
complement (20241..21420)
/tag= v
/note= "DH5"
complement (21420..22653)
/tag= w
/note= "AT5"
complement (22653..23892)
/tag= x
/note= "KS5"
complement (23992..46573)
/tag= y
/note= "fkBB gene"
complement (24163..24373)
/tag= z
/note= "ACP4"
complement (24997..26146)
/tag= aa
/note= "DH4 (inactive)"
complement (26146..27430)
/tag= ab
/note= "AT4"
complement (27430..28684)
/tag= ac
/note= "KS4"
complement (28750..28960)
/tag= ad
/note= "ACP3"
complement (29092..29740)
/tag= ae
/note= "KR3"
complement (29869..31018)
/tag= af
/note= "DH3 (inactive)"
FT
```


Db 62147 GCGCTCGGCTCCGCGGAGGACTTCGTGTCATGGCCCGCGATGGACCCCGGCACAGCCGAT 62206
Qy 1371 CAACGGGAGAACTGG---ATCGTCCGGAGCTGCGGCTCGCCCGACAGCCCTGGCGGCTG 1427
Db 62207 GACCGGCTCCGTACCGCCCATCTGAGCGGCTGCGAGGAGCGCGCGCTGCGCCG 62266
Qy 1428 GTACACCTCTCTGAGCACCCCGTGGCAGCGGTGACGAACCTACGTACCGGCACCAACAC 1487
Db 62267 TGCCGGGAGAGCTTCGCCCAAGCGCTGCGCGAGCTGCCCGACGCGCACCGCGGCGGC 62326
Qy 1488 ACTCCCGGACCGGACCTGCGAGCGGAGCGGCTGCTGCATGGACGGACGGCATACGA 1547
Db 62327 GGTGACCACTCTGCTCGGAGCGCCAGCGCGCGCTGCTCGGCCACGCGGACGCTCCGA 62386
Qy 1548 GATCGAGCTCGACATCGCTCGGACACCGGACGCAACGTCGGCATCTCGGTGGCGCGCTC 1607
Db 62387 GATCGCGCGGACCGAGCTTCAGGACCTCGGCATCGACTCGCTCACCGCGATCGAGCT 62446
Qy 1608 CCGGACGGAAACCGGCGCACAGACATCGGCAAGTACGGAGCAGACC---TGTAGTCTGA 1664
Db 62447 GCGCAACCGGCTCGCGGAGGCGGACCGGGCTGCGGCTGAGTCCACGCTGGTGTTCGACCA 62506
Qy 1665 CCGAGGACCTCCGACCTCGCGGGTACTCGCTCGCCCTTACTCGCGAGCGCGCGCC 1724
Db 62507 CCGGACCTCGGGTCTCGCGCGCCAGCTCGGACCGATCTGTTCGGACGCGCGTGCC 62566
Qy 1725 CATCGACCCCGCGCGCCGATCGGTGACCTGCGCATCTCGTCGACACCCAGAGCGTCTGA 1784
Db 62567 CACGCCCGCGGAGCGGACGACGACCCACGACGAGCCACTCGCGATCGTCGGCATGGC 62626
Qy 1785 GGTCTTCGTCAACCGCGCGCGCACCGTCTCTCCAGAGGTCTACTTCGCGCGAGGGCGA 1844
Db 62627 GTGCCGACTGCGCGCGGGGTGCGCTCGCCGAGGAGCCTGTGGCAGCTCGTGGCGTCCGG 62686
Qy 1845 CACGG 1849
Db 62687 CACCG 62691

Search completed: May 10, 2003, 06:29:26
Job time : 4644.5 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 04:44:16 ; Search time 2890 Seconds
(without alignments)
11213.553 Million cell updates/sec

Title: US-09-868-328B-2

Perfect score: 2001

Sequence: 1 gcggtgaccccgacttccc.....atcgacacgtctctcgctt 2001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_man:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.4	4.7	1288	14	BQ678719
C 2	90.6	4.5	1016	17	CNS03L09
C 3	88.2	4.4	1046	14	BQ952554
C 4	88	4.4	1651	12	BG809816
5	86.6	4.3	1065	14	BQ681076
6	86.6	4.3	1218	14	BQ876717

7	84.4	4.2	1339	13	BM458211
C 8	84.2	4.2	827	17	AZ183795
9	82.6	4.1	733	17	AG039579
C 10	82	4.1	934	17	AG046262
C 11	82	4.1	1023	17	AG128304
C 12	81.6	4.1	1341	17	AG030611
13	81.2	4.1	1242	14	BM911414
14	81.2	4.1	1281	12	BG852363
15	80.8	4.0	1051	14	BQ689718
C 16	80.6	4.0	1116	12	BG810038
C 17	80.6	4.0	1233	17	AG074706
C 18	80.4	4.0	1321	17	AG126084
19	79.8	4.0	1798	17	AG171124
C 20	79.4	4.0	1112	14	BQ058195
C 21	78.2	3.9	666	12	BG786337
22	78.2	3.9	724	17	AG171201
23	78.2	3.9	1061	17	AG046200
24	78.2	3.9	1150	14	BQ892481
25	78	3.9	775	17	AZ185516
26	78	3.9	821	17	AG063084
C 27	78	3.9	1243	14	BM913931
C 28	77.8	3.9	1143	10	AW731158
C 29	77.6	3.9	715	17	AG036665
30	77.6	3.9	1128	14	BQ710735
C 31	77.6	3.9	1170	17	AG111669
C 32	77.2	3.9	1297	14	BQ064843
C 33	77	3.8	1025	17	AG137560
C 34	77	3.8	1123	17	AG080476
C 35	77	3.8	1134	17	AG043044
C 36	77	3.8	1137	12	BG809979
37	76.8	3.8	1268	13	BM551006
C 38	76.6	3.8	1127	14	BQ718629
39	76.4	3.8	763	17	AG083422
C 40	76.4	3.8	1197	13	BI416470
C 41	76.2	3.8	1138	9	AI374006
42	76	3.8	1149	14	BM910848
C 43	76	3.8	1798	17	AG171124
C 44	75.8	3.8	1250	13	BM560215
C 45	75.8	3.8	1424	14	BQ720687

ALIGNMENTS

RESULT 1	BQ678719	1288 bp	mrna	linear	EST 15-JUL-2002
LOCUS	BQ678719	AGENCOURT_8199257	NIH_MGC_112	Homo sapiens	cdna clone IMAGE:6261563
DEFINITION	5', mRNA sequence.				
ACCESSION	BQ678719				
VERSION	BQ678719.1	GI:21791398			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 1288)				
AUTHORS	NTH-MGC http://mgc.nci.nih.gov/ .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabs-r@mail.nih.gov Tissue Procurement: DCTD/DTF CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCM2422 row: h column: 12 High quality sequence stop: 171. Location/Qualifiers 1. 1288				
FEATURES	source				

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/IMAGE="6261563"  
/clone_lib="NIH_MGC_112"  
/tissue_type="melanotic melanoma, cell line"  
/lab_host="DHL10B (phage-resistant)"  
/note="organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGGAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH_MGC Library."  
BASE COUNT      18 a      707 c      367 g      82 t      114 others  
ORIGIN  
Query Match      4.7%; Score 94.4; DB 14; Length 1288;  
Best Local Similarity 43.5%; Pred. No. 1.5e-06;  
Matches 394; Conservative 0; Mismatches 503; Indels 8; Gaps 3;  
QY 7 CACCCGAGCTTCCTCTCGACGACACCGTCCCTACCGCGCGAGACCGCGCCCGCCGACATG 66  
DB 236 CCNNCCCCCNCNNCCCCCGCCGCCNCCNCCCCCCCCCCCCCCTCCCTCCCGCCCC 295  
QY 67 TCTCTCAGCCTAGACGGGCGCCCTCTCTCGAGGTCTTCGTTCGGGACAGGTGAGGAGACTGG 126  
DB 296 CCGCGCGCCCGCCNCCCGCCCGCGCCCGCCCGCGCCCGCCCGCGCCCGCGCCGCGCN 355  
QY 127 TCGAAGCTGTCTCTCTGGGCGCGCGGTGTACCGCGCGAGCTCGAGCGCAGCGGCA 186  
DB 356 CGCGCCCGCCCGCCCGCCCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 415  
QY 187 GGAACCGTCAGTCAGCGCATGACGTCGAG-GCGGCCAGCGATGTCAGCGCCCTGA 245  
DB 416 CGCGGGGGCGCGCGCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 475  
QY 246 ACTTGGCGCGTTCCTGGGTGAGAGCGGTCCCGCCCGCCGACAGCTCTCTTCTACGG--C 303  
DB 476 CCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 535  
QY 304 TCGCCGACAGCGGTGAGCGCTTCGTGCGCGCCACCGCTCCAGAGAGAACAGCAATG 363  
DB 536 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 595  
QY 364 AGCGCGCGCATCTCAGCGCGCGCGCTGCTCCAGAGGAGCGCGCGCGCGCGCGCGCTG 423  
DB 596 NGCGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 655  
QY 424 ATCTTGGCGGTGTGTGCGCGCTGAGCGCGCGGATCCGCTCCCGGCTCGGTCTCGGTGCC 483  
DB 656 CCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 715  
QY 484 GTCTACACATGACGCGCGCGCGCGCTGCTGCGACCCCAAGCGCGCGCTGACACCG 543  
DB 716 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 775  
QY 544 CACGCGCGCTTACAGCTGTACTGTACTGTACTGTACTGTACTGTACTGTACTGTACTGT 603  
DB 776 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 835  
QY 604 GACACGCGAGCAGCAGCGCGCGCTGCGCTTCAGCAGCAGCGCGCGCGCGCGCGCGCG 658  
DB 836 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 895  
QY 659 CGCTGGGCGCGACTTCCCGGTGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 718  
DB 896 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 955  
QY 719 CAGGGTTCGCGCGCGCGCGGTGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 778  
DB 956 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015
```

```
QY 779 AGTACCAGGAGCAGTACTCTACTGTCGACGCGCGGGTTCAGTTTACCGCGCTGC 838  
DB 1016 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1075  
QY 839 CCGACCCCGCTCATCGTCAACACGACGCGTCTCGCGCGCGCCACACGCGCGCGAGATCGAGA 898  
DB 1076 GNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1135  
QY 899 AGGCC 903  
DB 1136 GCGCC 1140  
RESULT 2  
LOCUS      CNS03LU9/c      1016 bp      DNA      linear      GSS 17-MAY-2000  
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone  
036F09 of library G from Tetraodon nigroviridis, genomic survey  
sequence.  
ACCESSION  AL249930  
VERSION     AL249930.1      GI:7970942  
KEYWORDS    GSS: genome survey sequence.  
SOURCE      Tetraodon nigroviridis.  
ORGANISM    Tetraodon nigroviridis.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
REFERENCE   1 (bases 1 to 1016)  
AUTHORS     Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.  
TITLE        Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
JOURNAL      Unpublished  
REFERENCE   2 (bases 1 to 1016)  
AUTHORS     Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
TITLE        Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
JOURNAL      Unpublished  
AUTHORS     3 (bases 1 to 1016)  
TITLE        Genoscope.  
JOURNAL      Direct Submission  
AUTHORS     Submitted (12-APR-2000)  
COMMENT      This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/tetraodon.  
FEATURES  
source  
1..1016  
/organism="Tetraodon nigroviridis"  
/db_xref="taxon:99883"  
/clone="036F09"  
/clone_lib="G"  
/note="Genoscope sequence ID : C0B036CC05SP1-end :  
PUC-Ori"  
BASE COUNT      7 a      214 c      786 g      8 t      1 others  
ORIGIN  
Query Match      4.5%; Score 90.6; DB 17; Length 1016;  
Best Local Similarity 43.4%; Pred. No. 6.6e-06;  
Matches 414; Conservative 0; Mismatches 539; Indels 0; Gaps 0;  
QY 161 CCGCGAGCCTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 220  
DB 1012 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 953  
QY 221 CGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 280  
DB 952 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 893
```



```

Db 370 NCCCCCNCNNNNNNNNNNNNNNCTCTCTCCNNCNCNNNNNNNNNNNNNNNNNN 311
Qy 1853 TCTCGCTCTACACCGAGCGCGCGCCGACACACTTCCACGGGATCGTCTCCGCGAGATTG 1912
Db 310 CCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 251
Qy 1913 GCCAGCGGATCTAGGCGATGACACACACCGCTCTACCGAGCGCGCGCGGAGACGA 1972
Db 250 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 191
Qy 1973 CGGCGGACATCGACGCTCTCTGTCG 1999
Db 190 CCGTNNNNNNNNNACAGGNAACGGCG 164

RESULT 5
B0681076 1065 bp mRNA linear EST 15-JUL-2002
LOCUS B0681076
DEFINITION AGENCOURT_8187867 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6259803
5', mRNA sequence.
ACCESSION B0681076
VERSION B0681076.1 GI:21793755
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheraia; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2417 row: o column: 04
High quality sequence stop: 315.
Location/Qualifiers
1..1065
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6259803"
/clone_lib="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally-cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 63 a 593 c 314 g 55 t 40 others
ORIGIN
Query Match 4.3%; Score 86.6; DB 14; Length 1065;
Best Local Similarity 45.4%; Pred. No. 3.1e-05;
Matches 322; Conservative 0; Mismatches 381; Indels 6; Gaps 2;

Qy 9 CCGCGACTTCCTCGACGACACCGTCCCTACCGCGGACCGCGCGCGGCGGACTGCT 68
Db 307 CCGCGCGCGCGCGGAGACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 366
Qy 69 CCTAGCTAGACGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 128
Db 367 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426

```

```

Qy 129 GAACCTGCTCTCTCTGGGGCGCGCGGTGTGACCGCGAGCCTCGAGAGCGGACGCGCCAGG 188
Db 427 GCCCNGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 486
Qy 189 AACCGTGACGTGACCGCGGATCGAGTGTGAGGCGCGCGCGCGCGCGCGCGCGCGCG 248
Db 487 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 546
Qy 249 TGCGCGCGCTTCTGGGCTGACGAGCGCTCCACCGCGAGCTCTCTCTTCTACCGTGGT 308
Db 547 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 606
Qy 309 GAACCGAGGTGAGCGCTTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 368
Db 607 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 666
Qy 369 GGC----CATCTCAGCGCGCGCGCGTGTCTCAGGAGCGCGCGCGCGAGC--ACTCGCGCT 422
Db 667 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726
Qy 423 GATCTTCGCGCGTGTGTGCGCGCTGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 482
Db 727 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 786
Qy 483 CGCTTACACATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 542
Db 787 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 846
Qy 543 CCAGCGCGCGCTTACGAGTGTACTACTGCTGCTGCGCGCGCGCGCGCGCGCGCGCG 602
Db 847 CCNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 906
Qy 603 GGACCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 662
Db 907 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 966
Qy 663 GCGCGCGCGCGCTTCCCGCGTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 711
Db 967 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015

RESULT 6
B0876717 1218 bp mRNA linear EST 16-AUG-2002
LOCUS B0876717
DEFINITION AGENCOURT_8296665 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6249694
5', mRNA sequence.
ACCESSION B0876717
VERSION B0876717.1 GI:22268725
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheraia; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2391 row: i column: 23
High quality sequence stop: 308.
Location/Qualifiers
1..1218
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6249694"

... FEATURES
source

```

BASE COUNT 63 a 629 c 334 g 95 t 97 others
ORIGIN
/clone_lib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
Note: Organ: pancreas; Vector: pOTB7; Site: 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library.

Query Match 4.3%; Score 86.6; DB 14; Length 1218;
Best Local Similarity 41.9%; Pred. No. 3.1e-05;
Matches 334; Conservative 0; Mismatches 463; Indels 1; Gaps 1;
QY 7 CACCCGAGTTCCTCTGAGACGACGCTCCCTACCGCGCGAGCGCGCGCGCGACTG 66
DB 419 CCCCCGCGGCGCCNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 478
QY 67 CTCCTCAGCTAGACGGCGCCCTCTCGAGGTCTTCGTGCGGAGCGGTAGCGGACTGG 126
DB 479 CCCCNNCGCNCNCCCG 538
QY 127 TCGAACCTGCTCTCTGCGGCGCGCGGTGTGACCGCGAGCTCGAGCGGCGCGCGCA 186
DB 539 CCCCCG 598
QY 187 GGAACCGTGCAGTGACCGCGATGACGTGAGGGCGCCCGAGCGATGCTGACCGCCCTGAA 246
DB 599 CCG 658
QY 247 CCGCGCGCGGTCTGCGGTGAGAGCGGTCCACCGCGACACTCTCTTCTACGCGTGC 306
DB 659 CCGCGCNCNCCCG 718
QY 307 CGGAACAGGAGTGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366
DB 719 CG 778
QY 367 CGGGCATCTACGCGCGCGCGGTCTCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCG 426
DB 779 CCG 838
QY 427 TTCGCGGTGCTGTGCGCGGTGACGCGCGCGGTATCGGCTCGCGGTGCTGCTGCGGTG 486
DB 839 CGCNCGCGNCCCG 898
QY 487 TACCATAGACCG 546
DB 899 CCG 958
QY 547 GCGGCTTACCACTGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 606
DB 959 GCG 1018
QY 607 CACGCGAGCACGACGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 666
DB 1019 CG 1077
QY 667 CCGGACTTCCCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTG 726
DB 1078 CCG 1137
QY 727 GCGCGCGCGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTG 786
DB 1138 GCG 1197
QY 787 GAGAGTACTCTACTGG 804
DB 1198 GCG 1215

RESULT 7
LOCUS BM458211 1339 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6411296 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5530584
5', mRNA sequence.
ACCESSION BM458211
VERSION BM458211.1 GI:18507251
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1339)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM12210 row: o column: 01
High quality sequence stop: 134.
Location/Qualifiers
1. 1339
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5530584"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT 63 a 780 c 403 g 78 t 15 others
ORIGIN
Query Match 4.2%; Score 84.4; DB 13; Length 1339;
Best Local Similarity 47.9%; Pred. No. 7.4e-05;
Matches 414; Conservative 0; Mismatches 426; Indels 24; Gaps 5;
QY 2 CGGTGACCGCGACTTCCTCTGAGCAGCAGCAGCTCCCTACCGCGCGCGCGCGCGCGCG 61
DB 253 CGGCG 312
QY 62 GACTGCTCTCAGCCTAGACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 121
DB 313 GCG 372
QY 122 CTGCGCTGAACCTGTCTCTTGGGGGGCGGGGTGTGACCGCGAGCTCTGAGAGCGGAC 181
DB 373 CG 432
QY 182 GCGCAGGACCGTGCACGTGACCGCGATCGAGCTGAGCGCGCGCGCGCGCGCGCGCGCG 241
DB 433 GCG 492
QY 242 CTGACACTGCGCGCGCTTCTGGGCTGACGAGCGCTCCACCGCGCGCGCGCGCGCGCG 301
DB 493 CG 544
QY 302 GCTGCCCGAACCAGGGTGTGACGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 361
DB 545 GCG 604
QY 362 TGACCG 421


```
REFERENCE
AUTHORS      Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
              Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        BAC end sequences of Library PTB
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 733)
AUTHORS      Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
              Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
              1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
              (E-mail:chimbese@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
              Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT      Clones are derived from the chimpanzee BAC library PTB This BAC end
              was generated during the R&D process and may have higher chance of
              clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector       : pKS145
R.Site 1    : SacI
R.Site 2    : SacI
FEATURES
source      Location/Qualifiers
1. 733
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-016L16.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT  30 a 363 c 310 g 17 t 13 others
ORIGIN
Query Match 4.1%; Score 82.6; DB 17; Length 733;
Best Local Similarity 48.3%; Pred No. 0.00014;
Matches 287; Conservative 0; Mismatches 299; Indels 8; Gaps 2;
QY 157 GTGACCGGAGCTCTCTCTACCGCTGC--CCGAACAGGCTGACGCTTC-----G 328
Db 140 GCGACACCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 199
QY 217 GAGCGCCAGGATGCTGACGCCCTGAACCTGCGCGCGCTTCTGGGCTGACAGCGCTC 276
Db 200 GNGGGGGTGGGGAAGAACCGCGCCCTCCCTTGGCGCCCGCGCGCGCGCGCGCGCG 259
QY 277 CCACCCCGACAGCTCTCTCTACCGCTGC--CCGAACAGGCTGACGCTTC-----G 328
Db 260 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 319
QY 329 TCGCGCCACCGCTCCACGAGAGGAACAGCAATGACGCGCGCCATCTCACCGCGCGCG 388
Db 320 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 379
QY 389 TGCTCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 448
Db 380 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 439
QY 449 CAGCGCGGCGATCGCTCGCGGCTCGCTCGCTGCTTACCATGATGACGCGCGCGCGCG 508
Db 440 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 499
QY 509 GCTGGCTTGGACCGCCCAACCGCGCGGTACACACCGAGCGCGCTTACAGCTGTACTACC 568
Db 500 GCGCAGCGGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 559
QY 569 TCGACTCGACAGAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 628
Db 560 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 619
QY 629 TCGCGCTTACGACACGACCGGATGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 688
Db 620 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 679
```

```
QY 689 GGTCCGCGGCTCGTCGGCACCGCGAACACGCGAGGTTCCGCGCGCGCGCGGTGC 742
Db 680 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 733

RESULT 10
AG046262/c
LOCUS      AG046262
DEFINITION Pan troglodytes DNA, clone: PTB-025G21.R, genomic survey sequence.
ACCESSION  AG046262
VERSION    AG046262.1 GI:15583154
KEYWORDS   GSS.
SOURCE     Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
           BAC Library clone:PTB-025G21.R.
ORGANISM  Pan troglodytes
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE  1
AUTHORS    Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
           Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      BAC end sequences of Library PTB
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 934)
AUTHORS    Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
           Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      Direct Submission
JOURNAL    Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
           and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
           1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
           (E-mail:chimbese@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
           Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT    Clones are derived from the chimpanzee BAC library PTB This BAC end
           was generated during the R&D process and may have higher chance of
           clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector      : pKS145
R.Site 1   : SacI
R.Site 2   : SacI
FEATURES
source      Location/Qualifiers
1. 934
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-025G21.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT  30 a 201 c 654 g 16 t 33 others
ORIGIN
```

```
Query Match 4.1%; Score 82; DB 17; Length 934;
Best Local Similarity 44.8%; Pred No. 0.00018;
Matches 301; Conservative 0; Mismatches 369; Indels 2; Gaps 1;
QY 276 CCCACCGCGACAGCTCTCTCTTACCGTGCCTCCGACACAGGAGTGGAGCTCTCGTCGCGCC 335
Db 917 CCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 858
QY 336 CACCGCTCCACGAGAGAACAGCAATGACGCGCGCGCATCTCACGCGCGCGCGCTGTCCA 395
Db 857 NCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 798
QY 396 GGGAGCGCGCGCGCGAGGACTCGCCCTGATCTTCGGCGGTGCTGCGCGCTGCAGCGCG 455
Db 797 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 738
QY 456 GGCATCCGCTCGCGGCTCGCTCCGCTGCTTACCATGATGACGCGCGCGCGCGCGCGCG 515
Db 737 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 678
QY 516 CTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 575
```


REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : pKS145
R Site 1 : SacI
R Site 2 : SacI

Location/Qualifiers

1. 1341

/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-003A20.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 48 a 476 c 700 g 14 t 103 others

ORIGIN

Query Match 4.1%; Score 81.6; DB 17; Length 1341;
Best Local Similarity 44.8%; Pred. No. 0.00022;
Matches 388; Conservative 0; Mismatches 467; Indels 11; Gaps 5;

QY 36 CCCCCTACCGGCGGACCGCCGCGCGGCTGCTCCCTACGCTAGAGGCGCCCTCTCTCGA 95
DB 1299 CGCGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1240

QY 96 GTGCTTCTGCGGGAGCGGTGAGCGACTCGGTGGAACCTGGTCTCTCTGGGGCGGCGG 155
DB 1239 GCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1180

QY 156 TGTGACCGGCGGCTCGAGAC---GGCAGCGCGGAGACCGTGCACGTGACCGGATCG 211
DB 1179 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1120

QY 212 AGCTGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 271
DB 1119 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1060

QY 272 CGCTCCACCGCGACAGCT-CTCTCTTCTACCGCTGCGGCGGCGGCGGCGGCGGCGG 330
DB 1059 CGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1000

QY 331 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 390
DB 999 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 940

QY 391 CTCTAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 450
DB 939 -NCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 881

QY 451 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 510
DB 880 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 821

QY 511 TGGCTCTGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 570

Db 820 GNGNGNNGGCGG 761

QY 571 CACTCGGACGACAGAA---CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 626

Db 760 CGGNGGCGG 701

QY 627 CGTGGCTTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 686

Db 700 NNGCGG 641

QY 687 CGGCTCGG 746

Db 640 GCGG 581

QY 747 GC-TCCGACCGG 805

Db 580 GCGG 521

QY 806 CGACGACGCGCGGTTTACGTTTACGTTTACGTTTACGTTTACGTTTACGTTTACGTT 865

Db 520 GCGG 461

QY 865 GTGCGG 891

Db 460 GCGG 435

RESULT 13

LOCUS BM911414

DEFINITION AGENCOURT_6615160 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5466186 5', mRNA sequence.

ACCESSION BM911414

VERSION BM911414.1 GI:19361793

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1242)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLCM1966 row: c column: 19
High quality sequence stop: 350.

FEATURES

source

1. 1242

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5466186"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site:1: xhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 47 a 641 c 440 g 55 t 59 others

ORIGIN

```
Query Match 4.1%; Score 81.2; DB 14; Length 1242;
Best Local Similarity 43.5%; Pred. No. 0.00025;
Matches 373; Conservative 0; Mismatches 480; Indels 5; Gaps 3;

Qy 43 CCGCGGACCGCCGCGGCTCTCTAGCTAGACGGGCCCCCTCTCGAGTCTTC 102
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 274 CCGGAGGAGTTGGAGTGTGTGACCTGACGATGATGGGGTCCCGCAGGGCTGC 333
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 103 GTCGGGAGCGGTGAGCGACTCGCTCGAACCTGTGTCCTGCGGGGCGCGGTGTGACC 162
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 334 AGGAGGAGGNNNNNNNNNNCCNNNNNNNNCCCGCCCGCCCGCCCGCCCGCC 393
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 153 G-CGAGCTTCAGACGGCAGCGCAGGAACTGTGACGTGACCGGATGACGTGAGGC 221
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 394 GCGGCGCGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 453
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 222 GCCAGCGATGCTGACGCGCCCTGACCTGCGCGCGCTGTGGCTGACGAGCGTCCAC 281
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 454 CCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 282 CCGACAGTCTCTTCTTACCCTGCGCGGACAGGGTGGAGCTTCTGCGGCGCCACCG 341
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 514 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 573
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 342 TCCAGGAGGAAACAGCAATGACGCGCGGCAATCTACGCGCGCGCGCGCGCGGAGC 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 574 GCGCGGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 633
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 402 CGGCGCGGAGCACTCGCGCTGTCTTCTGCGGCGTCTGTGCGGCTGTGACGCGGGCATC 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 634 CCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 462 CGCTCCGCGCTCTCTCTTCCCTGCTTACACATGACGCGCGCGCGCGCGCTGTGCGA 521
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 694 GCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 522 CCGCGAGCGCGGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 581
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 754 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 582 GAACACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT--TCACG 639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 814 CNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 640 CACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 699
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 874 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 700 GTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 757
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 934 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 758 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 817
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 994 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1053
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 818 GTTACGTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1054 CNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1113
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 878 CCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 895
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1114 CNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 14
BG852363
LOCUS 1024034A03.y2 C. reinhardtii CC-1690, mRNA linear EST 29-MAY-2001
DEFINITION Chlamydomonas reinhardtii CC-1690, normalized, Lambda zap II
BG852363
ACCESSION BG852363
VERSION BG852363.1 GI:14233547
KEYWORDS EST.

Chlamydomonas reinhardtii.
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.
1 (bases 1 to 1281)
Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,
McDermott, J. P., Silflow, C., Stern, D., and Surzycki, R.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 2
Unpublished (2000)
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
Location/Qualifiers
1. 1281
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda zap
II"
/note="vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light, TAP medium in the dark, HS (minimal) medium in
ambient levels of CO2 and HS medium bubbled with 5% CO2.
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with ExAssist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT 30 a 697 c 456 g 15 t 83 others
ORIGIN

Query Match 4.1%; Score 81.2; DB 12; Length 1281;
Best Local Similarity 43.1%; Pred. No. 0.00026;
Matches 418; Conservative 0; Mismatches 547; Indels 5; Gaps 3;
Qy 9 CCGCGACTTCCTCTGACGACACACCTCCCTACCGCGCGCGCGCGCGCGCGCGCT 68
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 314 CCG 373
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 69 CCGCGACTGACGCTC 128
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 374 CCG 433
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 129 GAACCTGTCTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 188
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 434 GCG 493
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 189 AACCGTGCACGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 248
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 494 GCG 553
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 249 TCGCGCGCGTCTGCGCGTACGAGCGCTCCACCGCGCGCGCGCGCGCGCGCGCG 308
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 554 CCG 613
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 309 GAACGAGGTGAGCGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 368
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 614 GCGCG--GCG 671
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 369 GCGCATCTACGCGCGCGCGTCTCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCT 428
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 672 GCG 730
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```
QY 429 CGCGGCTGTGCTCCGCTGACGCGCGGATCCGCTCGGGCTCGCTCCGTCGCTCTA 488
Db 731 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 790
QY 489 CCACATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 548
Db 791 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 850
QY 549 CGCCTACACGCTGTACTACCTCCGACGAGAACACCGCGCGCGCGCGCGCGCG 608
Db 851 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 910
QY 609 CCGGAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 668
Db 911 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970
QY 669 CGACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 728
Db 971 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1030
QY 729 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 788
Db 1031 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1088
QY 789 GCGACTACCTACTGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 848
Db 1089 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1148
QY 849 CATCTGTCAACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 908
Db 1149 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1208
QY 909 GTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 968
Db 1209 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1268
QY 969 AGCACTCGCG 978
Db 1269 CCGCAGCGCG 1278

RESULT 15
BO689718
LOCUS BO689718
DEFINITION AGENCOURT_8344784 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6247838
5', mRNA sequence.
ACCESSION BO689718
VERSION BO689718.1 GI:21815034
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1051)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs_r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2386 row: 1 column: 15
High quality sequence start: 2
High quality sequence stop: 283.
Location/Qualifiers
1..1051
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```
/clone="IMAGE:6247838"
/clone_lib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: pancreas; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 63 a 557 c 291 g 69 t 71 others
ORIGIN
Query Match 4.0%; Score 80.8; DB 14; Length 1051;
Best Local Similarity 42.8%; Pred. No. 0.00029;
Matches 316; Conservative 0; Mismatches 415; Indels 8; Gaps 1;
...
QY 7 CACCCTGCTTCCCTCGACGACGCGCTCCCTTACCGCGCGCGCGCGCGCGCGCGCG 66
Db 310 CACCTGCTTCCCTCGACGACGCGCTCCCTTACCGCGCGCGCGCGCGCGCGCGCG 369
QY 67 CTCCTCAGCCTAGAGCGCGCGCTCTCTCGAGGCTTTCGTCGGGAGCGGTGAGCGACTG 126
Db 370 CCGCTCAGCCTAGAGCGCGCGCTCTCTCGAGGCTTTCGTCGGGAGCGGTGAGCGACTG 429
QY 127 TCGAACCCTGCTCTCTCTGCGGCGCGCGGTGTGACCGCGCGCGCTCGAGCGCGCGCG 186
Db 430 CCGAACCCTGCTCTCTCTGCGGCGCGCGGTGTGACCGCGCGCGCTCGAGCGCGCGCG 489
QY 187 GGAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 246
Db 490 GGAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 549
QY 247 CCGTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 306
Db 550 CCGTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 609
QY 307 CCGAACCAGGTGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 366
Db 610 CCGAACCAGGTGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 669
QY 367 CCGGCGCATCTACGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 426
Db 670 CCGGCGCATCTACGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 729
QY 427 TTCGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 478
Db 730 TTCGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 789
QY 479 GTGCGCGTCTTACCATATGACGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 538
Db 790 GTGCGCGTCTTACCATATGACGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 849
QY 539 CCACCCAGCGCGCGCTTACGAGTGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 598
Db 850 CCACCCAGCGCGCGCTTACGAGTGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 909
QY 599 GCTGGGACACGCGAGCAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 658
Db 910 GCTGGGACACGCGAGCAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 969
QY 659 CGCTCGCGCGCGCGCTTCCCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 718
Db 970 CGCTCGCGCGCGCGCTTCCCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1029
QY 719 CAGGCTTCGGCGCGCGCGCG 737
Db 1030 CAGGCTTCGGCGCGCGCGCGCG 1048
```

Search completed: May 10, 2003, 09:52:15

Job time : 2912 secs

THIS PAGE BLANK (USPTO)

QY 311 ACCAGGTGGAGCTTCTGTCGGGCCCCACCCGTCACAGAGAGAACCAACCAATGAGCCGG 370
Db 3941372 GTGCGCGCGGTGCTGCGGCTGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGG 3941313
QY 371 CCATCTACGCGCGCGCTGCTCCAGGAGAGCGGCGCGGAGCAGCTCGCCCTGATCTTCG 430
Db 3941312 CCGCGCGCGCGCGAGACACCGCGGCTTGCCTGTGTGTCGCGATCATGCGCGCGGCTTCG 3941253
QY 431 GCGGTGCTGTGCCGCTTCCAGCCCGGGGATCCGCTTCCGGGTTCGCTCGCTGCTTACC 490
Db 3941252 GTGCGCGCGCTGCG 3941193
QY 491 ACATGACGCGCGCGCGCTGCTGCGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 550
Db 3941192 CCGGTGCGCGCGGATGCGCGGTGCGGTGATGCGCGCGCGCGCGCGCGCGCGCGCG 3941133
QY 551 CTTACAGCTGTACTTACCTGCACTCCGACCAAGAACAGCGCGCGCGCGCGCGCGCGCG 610
Db 3941132 CTTGTCCGCGGTGCTA 3941073
QY 611 CGAGCACGACGAGCGGTGCGCTTACGACACCGAGCGCGCGCGCGCGCGCGCGCGCGCG 670
Db 3941072 CCGCGCGCGCGCGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941013
QY 671 ACTTCCCGTGTGTCGGGTGCGGCTGCTGCGGACCGCGCGCGCGCGCGCGCGCGCGCG 730
Db 3941012 CCGGACACGCTGTATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940953
QY 731 CCGCGCGCGCTGCTGCGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 790
Db 3940952 CCGCGGTGCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940893
QY 791 AGTACCTTACTGTGACGAGCGCGGTTCAGGTTCACCGCGCGCGCGCGCGCGCGCGCG 850
Db 3940892 CTTTGGCG 3940839
QY 851 TCGTCAACACGAGGTGCG 910
Db 3940838 CCGGTGCTTCCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940779
QY 911 TCCGCGACCCCAAGATCCACTGGGACACCGCGCGGAGATGGCTGCTGCTCATCGGAC 970
Db 3940778 CCGCGGTACCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940719
QY 971 GACTCGGTAGCGCGGTTCACACTTCGCGGAACTTCGCGGACGTGACACTTCGCGCGCA 1030
Db 3940718 CCGCGCGCGCGCGCTTGTGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCG 3940659
QY 1031 ACTTCACTACCGAACCG 1090
Db 3940658 GCGCGCGCGGAGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940599
QY 1091 CCGGAGACGCGGACACCG 1150
Db 3940598 CCGGCTTACCGGGAACCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940539
QY 1151 GCTTCCCATGACGTACCGCTTACTGACAGGCACTTGGAGCGGCGGAGAGTTCACGCG 1210
Db 3940538 GCGCGCGCGGTGCGCGGTACCGCGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCG 3940479
QY 1211 AGCACTTACCG 1270
Db 3940478 GCGCGCTCCGCTGGCGTGAAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940419
QY 1271 CATGATCGAGCGCGCGGAGACCAAGCGCTTCGCGATCGCGGTGATGAACACTGGAAGT 1330
Db 3940418 CTACCGCGGTGCGCGGTACCGCGCGGTGCGCGCGGTGCGCGGTGCGCGGTGCG 3940359
QY 1331 AGCGCGACGCGAGCTTCCCGACCGAGCGATCCGAGCGGTACACGCGGCGAGACTGATPC 1390
Db 3940358 GCGCGCGCGACCG 3940299
QY 1391 TCCGCGAGCTGCG 1450

Db 3940298 CCAGCGCCCGAGCAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940239
QY 1451 TGGAGCGCTGACCAACTA-----CGTACCGCGCGCGCGCGCGCGCGCGCGCGCG 1501
Db 3940238 GTCCGCGCGGTACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940179
QY 1502 CCGTGCAGCGGAGCGCGCTTCTGCGATGGAAGCGGCGCGCGCGCGCGCGCGCGCGCG 1561
Db 3940178 CCG 3940119
QY 1562 TCGCTGGGACACCGCGGAGCGTCCGATCTCGGTGCGCGCGCGCGCGCGCGCGCGCG 1621
Db 3940118 TGTCCGCGGTGCGCGCGGTAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940059
QY 1622 GGCACAC-----GAACATCTGCGAGTACGAGGAGCAAGCTGTACGTCCAGCGAGCGCT 1675
Db 3940058 CCGCGCGCGAGCAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3939999
QY 1676 CCGAGCTCGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1735
Db 3939998 CCGAGACCG 3939939
QY 1736 GCGCGCGATCGTCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1795
Db 3939938 GTCCG 3939879
QY 1796 AC CG 1855
Db 3939878 GCG 3939819
QY 1856 CGCTTACACGAGCG 1915
Db 3939818 CCCACAAATCGCGCG--GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3939761
QY 1916 AGCGATCTAGCGATGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1966
Db 3939760 CCGCGCGCGCGTTCGCGAGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3939710

RESULT 2
US-09-025-691-4
; Sequence 4, Application US/09025691
; Patent No. 6069299
; GENERAL INFORMATION:
; APPLICANT: Broadway, Roxanne M.
; APPLICANT: Harman, Gary E.
; TITLE OF INVENTION: FUNGUS AND INSECT CONTROL WITH
; TITLE OF INVENTION: CHITINOLYTIC ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/025,691
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/20120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600

Db 3936308 CGTCACCGCTGGCGCGCGCTGCTAGCCCGGTGCTGTTAGCGCCGCCCTTTCGCGCCTTGGC 3936249
Qy 195 GCACGTGACCGCGATGCGTGGAGCGCCAGCGATGCTACGCGCCCTGAACCTGCGCGC 254
Db 3936248 CGCGTGGCGCGCGACACCTTGGCGCGCTAAGTCTGGGTGCTCCCGTGGCGCGCG 3936189
Qy 255 GCTTGTGGGTGACAGAGCGCTGCC---ACCGGACAGTCTCTTCTTACCGGTGCCGAA 311
Db 3936188 CGCGCGCGCGCGCGCGCGCGCGGTGCGCGATCCGCGCTGCGCGGAAACACCG 3936129
Qy 312 CCAGGTGGAGCGCTGCTGCGCGCCACCGCTCCAGGAGAACAGCAATGACGCGCGC 371
Db 3936128 CGCGAATCCGGGTGCTCGCGCGCGCGCGCGATGCGCGCTTGGCGCGCTACCGCCTT 3936069
Qy 372 CATCTACGCGCGCGCGCTGCTCCAGGAGCGCGCGCGAGCACTGCGCCCTGATCTTCCG 431
Db 3936068 GCGCGCGCGCGCGCGCTGCTGCGCACGCTGCTGCTTCTGCGCGCTACACACCG 3936009
Qy 432 CGGTGCTGCGCGCTGACGCGCGCGGCAATCGCTCCGCGGTGCTGCTGCGCTTACCA 491
Db 3936008 CCGCGCGGTGCGCGCGCTGCGCGCGCGCGCTGATGCGCGCGCGCGGTGCGCG 3935949
Qy 492 CATGAGCGCGCGCGCGCTGCTGCGACCGCGCGCGCGCGCTGCGCGCGCGCGCGCG 551
Db 3935948 CGACCGCGCTTGGCGCGCTTGGCGCGCGCGCGCGAGCGCTTGGCGCTTGGAGAGCGCG 3935889
Qy 552 CTACAGCTGTACTACTCTCGAGCAGAACAAACCGCGCGCGCGCGCTGGGACACCG 611
Db 3935888 CGGTGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTC 3935829
Qy 612 GAGCAGACCGCGCGCTGCGCTTACGACACCGCGCGCGCGCTGCGCTGCGCGCGCG 671
Db 3935828 CCGCGTTCGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935769
Qy 672 CTTCCCGTGTGCTGCGCGTGGCGGTGCTGCGACCGCGCGCGCGCGCGCGCGCGCG 731
Db 3935768 TGGCG 3935709
Qy 732 CGGCGCGCTGCTGCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 791
Db 3935708 CCGCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935649
Qy 792 GTACTGTACTGTGCG 851
Db 3935648 CTTTGGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935592
Qy 852 GGTCAACACCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 911
Db 3935591 GGTCCCGTGGCG 3935532
Qy 912 CCGGACCG 971
Db 3935531 CGTGGCG 3935472
Qy 972 ACTGGGTACCGCGCGTCTACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1031
Db 3935471 GCGCGCGCTGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935412
Qy 1032 CTTGACTACCG 1091
Db 3935411 TCTGGCGTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935352
Qy 1092 CGGAGACCG 1151
Db 3935351 CGCGCGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935292
Qy 1152 CTTCCCATGACGTACGCGCTACTGACAGCGCGCGCGCGCGCGCGCGCGCGCG 1211
Db 3935291 CGTCTTGGGAGAGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935232
Qy 1212 CGACTCTACCG 1271

Db 3935231 CGCGGATACCGCGCTTGGCGCGCGTTCGCGCGGTGTTGGCGAGCGCGCGCGCG 3935172
Qy 1272 ATCGATGACGCGCGCGCGAGACCGCTCGCCATCGCGTGGATGAACAACTGGAAGTA 1331
Db 3935171 CGCGCGCGTCCCGCGGTGCGCGCGAGCGCGGATAAAGCGTGCCTGCGCGCGCG 3935112
Qy 1332 CGCGCGCGGACGTCCCGACCGAGCGATCGACGGCTACACGGGCGAGAACTCGATCGT 1391
Db 3935111 TGCGCGCGCGCGCGCGCGCGCGGTGCGCGCGT-----ACCG 3935071
Qy 1392 CGCGGAGCTGGCGCTGCGCGCGAGACCGCTGGCGGTGGTACACCTCTGAGCACCGCG 1451
Db 3935070 CTGGCGCGGTGGCTAGCGCGGTGCTGTTAGGCGCGCTTTGGCGCGCTTG--CGCGCT 3935013
Qy 1452 GGCAGCGTACGAAGTACGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1511
Db 3935012 GCG 3934953
Qy 1512 CAGCGCGCTGCTGCGATGGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1571
Db 3934952 CCG 3934905
Qy 1572 CACCGCGAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1631
Db 3934904 CACCG 3934845
Qy 1632 CATCGCGAAGTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1691
Db 3934844 ATCGCG---GTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934788
Qy 1692 CTGCGTCCCGCGCTTACTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1751
Db 3934787 GTCGCGCGATCAGGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934728
Qy 1752 CTTGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1811
Db 3934727 CGCGCGCGCGCGTGTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934668
Qy 1812 GCTCTCCCGAGGTGCTTCTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1871
Db 3934667 GCGCGCGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934608
Qy 1872 CG 1931
Db 3934607 CG 3934548
Qy 1932 GCACACCGCGCGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1962
Db 3934547 CACCG 3934517

RESULT 4

US-08-403-852D-1

; Sequence 1, Application US/08403852D

; Patent No. 5891695

; GENERAL INFORMATION:

; APPLICANT: Blanc, Veronique

; APPLICANT: Blanche, Francis

; APPLICANT: Crouzet, Joel

; APPLICANT: Jacques, Nathalie

; APPLICANT: Lacroix, Patricia

; APPLICANT: Thibaut, Denis

; APPLICANT: Zagorec, Monique

; APPLICANT: Debussche, Laurent

; APPLICANT: De Crecy-Lagard, Valerie

; TITLE OF INVENTION: Polypeptides Involved In The

; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences

; TITLE OF INVENTION: Coding For These Polypeptides And Their Use

; NUMBER OF SEQUENCES: 43

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner

; STREET: 1300 I Street, N.W., Suite 700

; CITY: Washington

STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852D
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PC7/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806, 0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaeaspiralis
US-08-403-852D-1

Query Match 4.6%; Score 91.4; DB 2; Length 5392;
Best Local Similarity 42.1%; Pred. No. 1.8e-08;
Matches 652; Conservative 0; Mismatches 891; Indels 6; Gaps 2;

Qy 382 CGCGCGGTGCTCCAGGAGCGCGCGGAGCACTCGCCCTGATCTTCGGCGGTGCTGTG 441
Db 2748 CGGACAGATACCCGCGCGGTGCGGCGCGGAGCGCCGGATTCGAGCGCGTCTGTAT 2807

Qy 442 CGCGCTGAGCGCGCGGATCGCGTCCGGCTCGCTCCGCTGCTACACATGACGCGCC 501
Db 2808 CGACGACGCGCGCGCGCGGTCCAGGCGGTTCAGAGGAGCAC--GCTGACCGCC 2865

Qy 502 CCCAGCGGTGCTGCTGGAGCCCCAACGCCCGGTTCACACCCAGCGGCGCTTACAGCTG 561
Db 2866 GCGCTGGCGCGGTCAACGAGCACATCGGCTGTATCACCGCCCGCTCCGCGCGGACCA 2925

Qy 562 TACTACCTGCACTCCGACAGACAGCGCGCGCGCGGTGGAGCAGCGAGCAGCACC 621
Db 2926 GCGCGCTACCAAGCTGTCGCGATCACCGCTCGCTCGACACCTCGCGCGCGCGCGC 2985

Qy 622 GACGCGGTGCGCTTCACGACACCGACCGCTGTGCTGCGCTGCGGCGCGCTTCGCCG 681
Db 2986 GCGTGGCTCGGAGCAGCGACACCGACCGCGGCGCGGCGCGGCGGCGGCGGCGG 3045

Qy 682 TGGTCCGCGGTGCGGTGCGGACCGCGGAGCGGATTCGGCGCGCGCGCGGCGGTC 741
Db 3046 GTCGTCCGCGCGGTGCGGAGAGCTTCGACGAGCGGCTTCGTCCAGCAGCGCGCGC 3105

Qy 742 GTCGCGGTGCGGACCGCGGAGCGCGGTTCGCGAGTACAGGAGCAGTACCTCTAC 801
Db 3106 GCGCTGTACTGCGGCGGTGCGCGCGGTCCACCACTCGACACCGCGGCGGCGGCGG 3165

Qy 802 TGGTACCGAGCGGCGGTTCAGCTTCACGCGCGCTGCGCGCGCTCGTACGTCAGCACC 861
Db 3166 GTGGCGCGCGCGCTCAAGCTGCGCGCGCGCGCGCGGCGCGGCGCGCGGCGGCGG 3225

Qy 862 GACGCTCGCGCGCGGACACCGCGCGGAGATCGAGAACCGCGAGTGGTTCGCGGACCCC 921
Db 3226 GCGCGCGCGCTCGCGCGCGCGCGGACCTCGTCTGCTCGAGGCGCGCGCGCGCGCC 3285

Qy 922 AAGATCCACTGGGACACCGCGCGGAGAAATGGGTCTCGGTCTATCGGAGCGACTCGG 981
Db 3286 TCGGTGAAGCAGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 3345

Qy 982 GCGCGGTTCACACTCGCGCGGAACTCGCGGACTGGAGACTTCGCGCGCAACTTCGCTAC 1041
Db 3346 GAACTGCGCGCGGACCGCGCGCGCGCGCGGCTTCACGGTGGCGCTCACCGGTTCGAGC 3405

Qy 1042 CGGAACACGCGCGCGCGGATCGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1101
Db 3406 CCGGTCTGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3465

Qy 1102 GGGACACCGCACTGGGTCTCGCGCGCGAGATGGAGCGCTACGGATCGGCTCGCCCATG 1161
Db 3466 CTGGCGAAGCGCTGGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3525

Qy 1162 ACGTACGCTACTTGACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1221
Db 3526 CGACCGCTCGCGCGCGCTCTCTCTGGAGA---GTATGTCCCGTCTGCTTCACTCG 3581

Qy 1222 CGCAATGCTCGACTGGGCTGGAGCTGTACGGCGCGCGCGCGCGCGCGCGCGCGCG 1281
Db 3582 GAGTCCGTGACCGGCG 3641

Qy 1282 GCGCGCGAGACCGCGCGCTCGCCATCGGTGGATGACAACTGGAAGTACGCGCGCGC 1341
Db 3642 GACGCGCTGCTGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3701

Qy 1342 GACGTCCCG 1401
Db 3702 GCGCAGGTCCACATCG 3761

Qy 1402 CGGCTGCG 1461
Db 3762 GTCCGCGAGCGATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3821

Qy 1462 ACGAATACGTACCG 1521
Db 3822 TCCTGCGCGCGTCTCGGTCTCCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3881

Qy 1522 CTGCCATGGAAGCG 1581
Db 3882 AGCGCTACGAGACCG 3941

Qy 1582 AACGTGCG 1641
Db 3942 GACGAGGCGCTGATGTTGCGGTACGCCACCGCGCGCGCGCGCGCGCGCGCGCG 4001

Qy 1642 TACGAGCAGACCTGTAGCTCGACCGGAGACCGCTCCGCGCGCGCGCGCGCGCG 1701
Db 4002 ATCGAGCTCG 4061

Qy 1702 CCCTACTCG 1761
Db 4062 CCCTACTCG 4121

Qy 1762 CTCGTCGACACCG 1821
Db 4122 GTGCGCTGGACCGCTGCTCTCTCTCCGAGCAGCGCGCGCGCGCGCGCGCGCG 4181

Qy 1822 CAGGTTCACCTTCG 1881
Db 4182 CTGCTACCG 4241

Qy 1882 CACTTCACCG 1930
Db 4242 GACGCGATCAAGCTCGAGACGCGGACAACTACCGCGCTGCTGGTCAACCGCGA 4290


```

Qy 1674 CTCGCACTTCGGCGGACTCTCGTGCAGCCCTCTACTCGGAGCGCGGCCCATCGACCC 1733
      ||||| | | |||| | | |||| | | |||| | | |||| | | |||| | | ||||
Db 27843 GGCGGAGCGCGGAGCGCTCTGGGGGGCACCCCGCCCCCGCGCGGAACAACCTGGACAT 27902
      ||||| | | |||| | | |||| | | |||| | | |||| | | |||| | | ||||

Qy 1734 CGGCGCGGATCCGTGCACCTTGCCATCCTCGTCGACACCCAGAGC 1779
      .||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 27903 CGGCGACTTCTACCAGCGCTTCTCCGAACCTCGGTTACGGCTACGGC 27948
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kubstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R. A.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rosteck, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4437 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
;
US-08-804-198-1

Query Match 4.38; Score 85.6; DB 2; Length 44377;
Best Local Similarity 42.18; Pred. No. 2e-07; Indels 30; Gaps
Matches 727; Conservative 0; Mismatches 969;

Qy 69 CCTCAGCCTAGACGGGCCCCTCTCTCGAGGTCTTCTCGGGGACGCGTACGGCTCGGTC 128
      || ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db 26238 CCGCACCCTGTCGACGACGCGGCGCCCTGCTCGCGGAGAACTTCACGAGCTGATCCG 26297
QY 129 GAACCTGGTCTCTCGGGGCGGCGGTGTGACCGGAGACCTCGAGACGCGACGCGCAGG 188
Db 26298 GCGCTCGACGACCTGCTCGAAGGCGACCGACCGCTGCTGCGGGGCGACCGCGG 26357
QY 189 ACCGTGACGTGACCGCGATGACGTGAGAGGCGCCACCGGATGAGCGCCCTGAACC 248
Db 26358 CACCTCGACGAGTGTGCTGCTTCCCGCGCCAGGCTGCGAGTGGCGCGAGATGC 26417
QY 249 TCGCGCGCTGCTGGGCTGACGAGCGCTCCACCGACAGCTCTCTCTTACGGTGGCC 308
Db 26418 GCGAGGCTGCTGGCCCGCTCAGCGGCTCCGCGCTCTTCTTGGAGACCGCCCGCGCTG 26477
QY 309 GAACAGGTGACGCTTCTGTCGCGCCACCGCTCCACGAGAGAACACGCAATGACGCC 368
Db 26478 CGACCTCGCGCTCCGCGCCACCTCGGCTGGTGGTCTGAGCTACGCGCGGGAACC 26537
QY 369 GCGCATCTACGCGCGCGCTGCTCAGGAGCGCGCGGAGACACTGCGCCCTGATCTT 428
Db 26538 CCGCGCGCTCTCGTGGACCGGCTGACGTGTGACGCGCGCTGCTGTTTACCATGATGT 26597
QY 429 CCGCGGTGCTGTGCGCTGACGCGCGGCATCCGCTCGGGCTGCTCTCCGTGCGGTCTA 488
Db 26598 CTCGCTCGCGAGAGCTGGCGTTCGCTGGCGCTGACGCGCGCGGTGCTGCTGCTC 26657
QY 489 CCACATGACCGCGCGCGCTGCTGCGACCGCGCGCTGCGACCGCGCGCTGCGACCGG 548
Db 26658 CCAGGCGAGATCGCGCGCGCTACGTGCGCGCGCTGACGCTGAGACGCGCGCGG 26717
QY 549 CCGCTACGAGCTGTACTACTCTGACTCGACCGACGACGCGCGCGCGCGCGCGG 608
Db 26718 CATCTGCGCGCTGCGACGCGCGCTGCGTGGCGTGGCGGCAAGGCGCGATGTCG 26777
QY 609 CCGGACGACGCGCGCGCTGCGCTTACGACGACGCGCGCTGATGCGCTGCGCG 668
Db 26778 GGTGACCTGTGCGACGCGCGCTGCGTCCCGCTGAGCGCTGAGGCGCGCTGCG 26837
QY 669 CGACTTCCCGCTGTGCTGCGGCTGCGGCGTGTGCGACCGCGGAGACGCGAGGTT 725
Db 26838 CCGTGGCGCGCTCAAGGCGCGCGAGACCTGCGCGCTGCTCGGGGACCGCGCGCTGC 26897
QY 726 -----CGGCGCGCGCGGCTGCTGCGGCTGCGGCTGCGGCGCGCGCGCGGCT 773
Db 26898 GAGCTGTGCGCGACTGCTGCGGAGGCGGTGACGCGCGCGCGCGCGCGCGCTGCG 26957
QY 774 CCGCAAGTACGAGGACGCTACTCTACTGTGCGACGCGCGGCTTACGCTTACCGC 833
Db 26958 CACCGCGCGCGCTGCGCGCGAGTGCACGCTGAGGCGCGCGCTGCGGAGGCTGCTGC 27017
QY 834 CCGTGGCGCGCGCTGCTCAACACGCGCGCTGCGCGCGCGCGCGCGCGCGCGAGAT 893
Db 27018 GCGCGTGGCG 27077
QY 894 CGAGAACGCGCGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGATG 953
Db 27078 CGACACCGCGCGAGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGTT 27137
QY 954 GGTCTGCGTCACTGCGAGGCTGCGGTGACGCGCGCTTCTACACTCGCGAAGCTCGCGGA 1013
Db 27138 CGAGCAGCGCGCGCGCGCGCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 27197
QY 1014 CTTGGACACTTCCCGCGAACTTCGACTACCGCGAAGCGCGCGCGCGCGCGCGCGCG 1073
Db 27198 GCACCG 27257
QY 1074 CGACCTGTTCGAGATCACCGGAGACGCGGACGCGCGCGCGCGCGCGCGCGCGAGAT 1133
Db 27258 CGTCTCGCGACCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 27317
QY 1134 GCGCGCTACGCGCATCGCGCTCCCGCTACGTAAGTACGCTACTGACGAGGCGCGCGG 1193
Db 27318 CCGCGCGCTACACCGCGCGCTGAGATCGACCGCGCGCGCGCGCGCGCGCGCGCG 27377

QY 1194 CGAGCAGTTTCCAGCGCGAGACCTCACCCCGCAATGGCTGAGTGGGCTGGGACTGGTA 1253
Db 27378 CCAGGTGGAACCTGCCACGCTACCCCTTCAGCGCGAGCGCTACTGTACAGCCCGCGCA 27437
QY 1254 CCGCGCGCTGCTGCGGCTGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1313
Db 27438 CCGCGGTGACGACCGCGCGCTCCCTCGGTCTGGACCGCGCTGCGCGCGCGCGCTG 27497
QY 1314 GATGAACAACTGGAAGTACGCGCGCGAGCGCTCCCGACCGCGCGCTGCGCGCGCTG 1373
Db 27498 CCGCGTGCAGTCCCGGAGTCCGCTGATGCGGATGTACCGCGCGCGCGCGCGCG 27557
QY 1374 CCGCGCAGAACTCGATGCTCGCGAGCTGCGGCTGCGCGCGCGCGCGCGCGCGCTG 1433
Db 27558 CACCCGCTGCTGCGCGCGCGCGCTGCTGGGCTGCGCGCTGCTGCGCGCGCGCGCT 27617
QY 1434 CCGCTGAGCAGCGCGCGCGCGCGCTGACGAACTGCTGCGCGCGCGCGCGCGCGCT 1493
Db 27618 CG-----CGACCTGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 27668
QY 1494 CGACCGGACCGTGCAGCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCTG 1553
Db 27669 GAGCTCAGCTGCGCGCGCGCGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCTG 27728
QY 1554 GCTCAGACTGCGCTGCT 1613
Db 27729 GAACGTGCG 27788
QY 1614 CGGACCG 1673
Db 27789 GCG 27842
QY 1674 CTCG 1733
Db 27843 GCG 27902
QY 1734 CCG 1779
Db 27903 CCG 27948

RESULT 9

US-07-945-283-1/c
; Sequence 1, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EP0 and LLT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Peoria
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
; FILING DATE: 19920911
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ribando, Curtis P
; REGISTRATION NUMBER: 27976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 309-685-4011 ext.513

TELEFAX: 309-685-4128
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8438 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudorabies virus
FEATURE:
NAME/KEY: CDS
LOCATION: 622..6495
FEATURE:
NAME/KEY: variation
LOCATION: replace(1099, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1267, "t")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1381, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1566, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(7010, "g")
US-07-945-283-1

Query Match 4.2%; Score 84; DB 1; Length 8438;
Best Local Similarity 43.08; Pred. No. 3.7e-07;
Matches 640; Conservative 0; Mismatches 830; Indels 18; Gaps 4;

Qy 329 TCGGCGCCACCCGTCACGAGAGAACAGCAATGACGCCGGGCATCTACGCGCGCGCG 388
Dy 5549 TCG 5490
Qy 389 TGCTCCAGGAGCG 448
Dy 5489 CCTCCTCGCGCTCCG 5430
Qy 449 CAGCG 508
Dy 5429 CCG 5370
Qy 509 GCTGGCTGTGCGACCCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 568
Dy 5369 ACGGGCG 5310
Qy 569 TGCACTCCGACAGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 628
Dy 5309 CCGACCTTCCG 5250
Qy 629 TCGCGCTTACACACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 688
Dy 5249 TCTGGAGGAGGAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5190
Qy 689 GGTGCGCGGTGTGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 748
Dy 5189 TGCGCGCTTCTATCCGCGAGATGGGGACTCGAGGAGCAGCAGGAGCGCGCGCG 5130
Qy 749 TCGCGACCCACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 808
Dy 5129 TCATCTACAGCG 5070
Qy 809 CCGACGCGGGTTCACGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 868
Dy 5069 CCGCGGACCGCGCTTACACAGTTCTGCCAGCGCGCGCGCGCGCGCGCGCGCG 5010
Qy 869 GCG 928

RESULT 10

US-08-951-742-1
; Sequence 1, Application us/08951742
; Patent No. 6127144
; GENERAL INFORMATION:
; APPLICANT: Bartfield, Daniel
; APPLICANT: Michael J. Butler
; APPLICANT: Dany Hadary
; APPLICANT: David Jenish

Db 5009 GCTCCTTTCATCACCGGAGCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4950
Qy 929 ACTGGGACACCGCGCGCGGAGAAATGGGTCTCGGTATCGGAGCACTGGGTACCGCGGT 988
Dy 4949 CCCAGGACCGCGCTCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4890
Qy 989 TCTA---CACCTCGCGCAACTCGCGGACACTGCGGCGCAACTTCGCGCACTACCGCA 1045
Dy 4889 ACGATCGCACCCAGAACCTTCATCTCGAGAGCCTTCGCGCGCGCGCGCGCGCG 4830
Qy 1046 ACCACGCGCTCGCGGCGCATGAGTGCCTCGGACCTTGTTCGAGATCACCGAGACAGCGGA 1105
Dy 4829 CCTACCG 4770
Qy 1106 CACGCCACTGGTGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1165
Dy 4769 CCGCGCTCG 4710
Qy 1166 ACGCGT---ACTGGACAGCGCACTGGGACCGCGGAGCGAGTTCCACCGCGCGCGCGCG 1222
Dy 4709 ACGCTGCGTCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4650
Qy 1223 CGCAATGGCTCGACTGGGGCTGGGACTGGTACGCGCGCGCGCGCGCGCGCGCGCG 1282
Dy 4649 TGCG 4590
Qy 1283 CGCGCGGAGACCAAGCGC-----TCGCCATCGGTGGATGAACAACCTGGAAGTACG 1333
Dy 4589 CCGCGCTGCTGCGCTCGCGGACCGGATCCCGCGCGCGCGCGCGCGCGCGCGCGCG 4530
Qy 1334 CCGCACGCGAGCTCCCGACCGACGATCCGAGCGCTACACGCGGCGCAACTTCGATCGTCC 1393
Dy 4529 CCGCGGACTGTGTCGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4470
Qy 1394 GCGAGCTGCGGCTCG 1453
Dy 4469 ACGGGGCGCG---CGCGCGCGGAGCGCGCTGGACCTACGCGCGCGCGCGCGCG 4413
Qy 1454 CAGCGCTGAGCAACTAGTGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1513
Dy 4412 CCAAGTGGCG 4353
Qy 1514 GCGCGCTCTGCCATGGAACGAGCGCGCATACGAGATCGAGCTCGACATCGCTTGGGACA 1573
Dy 4352 GCGCGCGCTTGCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4293
Qy 1574 CCGCGAGAGCTCGGCACTCTGGTGGCGCGCTCCCGCGAGCGGAACCGCGCGCGCG 1633
Dy 4292 CCG 4233
Qy 1634 TCGCAAGTACGAGAGAGCTGTACGTGACGCGGAGGACCTCCGACCTCGCGGGTACT 1693
Dy 4232 CCAAGTCCGGTCCAGCACCAAGTCCAGCTCCGCGCGCGCGCGCGCGCGCGCG 4173
Qy 1694 CGTTCGCGCGCGCTACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1753
Dy 4172 CCGGCTACGCGAGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4113
Qy 1754 TCGCGCATCTCTGTCAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1801
Dy 4112 AGCG 4065

APPLICANT: Tim Krieger
 APPLICANT: Lawrence T. Malek
 APPLICANT: Gisela Soosmeyer
 APPLICANT: Eva Walezyk
 APPLICANT: Phyllis Krygsmann
 APPLICANT: Sheila Garven
 TITLE OF INVENTION: METHOD FOR EXPRESSION OF PROTEINS IN BACTERIAL HOST CELLS
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/951,742
 FILING DATE: 16-OCT-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Bent, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 0189740/0140
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1908 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 146..1756
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 146..148
 OTHER INFORMATION: /product= "Met at position -39"
 OTHER INFORMATION: represents fMet"
 FEATURE:
 NAME/KEY: sig_peptide
 LOCATION: 146..262
 FEATURE:
 NAME/KEY: mat_peptide
 LOCATION: 263..1756
 US-08-951-742-1

Query Match 4.1%; Score 83; DB 3; Length 1908;

Best Local Similarity 43.2%; Pred. No. 5.4e-07;
Matches 560; Conservative 0; Mismatches 720; Indels 15; Gaps 3;

QY 610 GCGAGCAGCAGCGGGGTCGCTTACGACACCGCAGCGGTCGATCGCGTGGCGCC 669
 DB 101 GCGAGATCCCGTACTTGTTCGGACACGTCACGGGAGGGCCACATGAGGAAGACG 160
 QY 670 GACTTCCCGGTGTGTGGGTTCGGGTTCGGACACCGGCAACGCGCAGGGTTCGGC 729
 DB 161 ATACGGCGGAGGCGGCGGTCGCGACGCGCGGAGCAGCTGGTACCGCCACGCTGATC 220
 QY 730 GCGCGCGGTGTGTGGGTTCGGGTTCGGGTTCGGGTTCGGGTTCGGGTTCGGGTTC 789
 DB 221 GCGCGCGGTGTGTGGGTTCGGGTTCGGGTTCGGGTTCGGGTTCGGGTTCGGGTTC 280
 QY 790 CAGTACCTCTACTGTCTGACGACGCGCGGGTTCAGTTTCACGCGGCTTCCCGACCCCGTC 849

DB 281 CGGAGCTGGACCGGAGCGCGGTGCCGCCATCGCCGCCGCCGCCGCCGCCGCCGCCG 340
 QY 850 ATCGTCAACACGAGCGGTGCGCGCCGCCACACCGCCGCCGAGATCGAGAACGCCGAGTGG 909
 DB 341 GGCATCGACTGGGAGGACTGCGCAGCGGACTGGAAC-----TGCCCAA;CCCATCCAG 394
 QY 910 TTCGCGGACCCCAAGATCCACTGGGACACCGCCCGGGAGATGGTCTGGCTCATCGGA 969
 DB 395 TCGGCTACGTACGGTCCGATGGACTACGCCCAAGCCGTACGGCAAGCAGATCAGGCTC 454
 QY 970 CGACTGCGGTACGCGCGTTCCTACACCTCGCGGAACCTCGCGGAGCTGGACACTTCGCGCG 1029
 DB 455 GCCGTCGACCGCATCGCAACACCGAAGAGGAGCGCGCGGCGCCCTCATCTAC 514
 QY 1030 AACTTCGACTACCCGAACACGCCCTCGCGGCGCATCGAGTCCCGGCGCTGTTCAGATC 1089
 DB 515 AACCCCGCGGTTCGCGCGGTTCGCGCTCGCTTTCCCGCGCGCGCTCACGAACAGAGC 574
 QY 1090 ACCGCGAGGAGGACAGCGGCTGGGTGTCGCGCGCGCAGGATGGAGCGCTTACGGCATC 1149
 DB 575 GCGGTCTGGGCCAACACGCGCAAGGCCCTACGACTTTCGTGGCTTCGACCGCGCGCGCTC 634
 QY 1150 GSCCTCCCATGACGTACGCTTACTGGACAGGCACTCGGACGGCGAGCAGTTCCACGCC 1209
 DB 635 GGCCTCTC---GCGCCCATCTCTCTGCGTGCACCGCAGGAGTTCGTCAGGCAACCCAAAG 691
 QY 1210 GACGACTCACCCCGCAATGGCTCGACTGGGCTGGGACTGGTACGCGCGCGCTCACCCTGG 1269
 DB 692 GCGGACCCCGTGGCGGCTCGGAGCGCCACAAAGCGCGCCAGCAAGTTCGCGCGCGAG 751
 QY 1270 CCATCGATCGAGCGCGCGGAGACCAAGCGCTCGCCATCGCGTGGATGAACAACTTGAAG 1329
 DB 752 TACGCGGAGGGCTGCTTCGAGCGCAGCGCGCGAGATGCTCCGCGCATACCCACGCGCAAC 811
 QY 1330 TACGCGCACGCGAGCTCCCGACGACATCCGAGCGCTACACGGCGAGAACTCGATC 1389
 DB 812 ACCGCGCGGACCTCGAGCTCATCGCGCGCGCTTCGCGGAGAGAACTCACT----- 866
 QY 1390 GTCCGCGAGCTGCGGCTCGCCCGACAGCCTGGCGGCTGGTACACCTCTCTAGCACCCCC 1449
 DB 867 -ACCTCGCGGCTCTCTACGCGACCTTACCTCGCGCGCTTACGCGCGCTCTTCGCGGAC 925
 QY 1450 GTGGCAGCGCTGACGAATAGTGTACCGCGCACCGACACCTCCCGCGCGGACCGCTCGAC 1509
 DB 926 CAGCTCCCGCGCATGGTCTGACAGCGTGTCAACCCGTCGCGGAGCAAGATCTGGTAC 985
 QY 1510 GGCAGCGCGCTCTGCCATGGAACGAGCGCATACGAGATCGAGCTCGACATCGCCTGG 1569
 DB 986 CAGGCCAACCTGGACAGGACGTGCGCTTCGAGGCGCGCTGGAAGGACTGGCAGGACTGG 1045
 QY 1570 GACACCGGAGAGCTCGGATTCGCGTGGCGCGCTCCCGGAGGAACTCCGGGACAGG 1629
 DB 1046 GTCCGCGGAGACGACGCGCGCTTACCACTCGCGCGGACACCGCGCGGAGTTCAGGAC 1105
 QY 1630 AACATCGGCAAGTAGGAGAGACCTGTACGTGACCGGAGGACCTCGACCTCCCGCGG 1689
 DB 1106 TGGCTGAAGCTGCGCGCGCGCGCGGAGAAAGCGCTGGCGGGGCTGTCGGACCGGCC 1165
 QY 1690 TACTCGCTCGCGCGCTTCTCGCGAGCGCGCGCGCTTACGACCGCGCGCGCGCGCTG 1749
 DB 1166 GAGCTGATCTCTTCTCCAGAGCGCGCGCTTACTACGACTCCGCTGGCGCGCGCGCG 1225
 QY 1750 CACTTCGCGATCTCGTTCGACACCGCAGAGCTCGAGGCTTCGTCAACGCGCGGCGCACAC 1809
 DB 1226 GAGATCTTTCAGCAAGTACGTGCGCGCGGACACCGCGCGGCTCGTGCAGCGCGCGC 1285
 QY 1810 GTGCTCTCCCGAGGCTCCACTTCGCGGAGGCGGACGCGGAATCTGCTCTACACCGAC 1869
 DB 1286 GACCTGTCCGACACCGCGGCAACGCTTCGCGGAGAGACGGAACCGCTCTACAGGCC 1345
 QY 1870 GCGCGCGCGCGCACACTTCACCGGATCTGTCGTCG 1904
 DB 1346 CTCGAGTGCACCGAGCGCAAGTGGCGCGCGCAACTG 1380

RESULT 11

US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 4.1%; Score 83; DB 4; Length 4403765;
Best Local Similarity 42.6%; Pred. No. 6.6e-07;
Matches 774; Conservative 1; Mismatches 1029; Indels 12; Gaps 4;
QY 86 CCTCTCGAGGTCTTCTGCGGACGCTGAGGCGACTGGGTGCAACTGGTCTCTCTGG 145
Db 1631220 CCTCTGTTGTCGAGGAGCGCCTTGGCGCGGTCTCGGCGCGCCAGTCCCGCG 1631279
QY 146 GGGCGCGGTGACCGCGAGCTCTGAGACGCGACGCGCGGACGAGTGCAGTGCACG 205
Db 1631280 ATGCCACCGTTCGCGCGGTGGCGCGGTGCGCGCTTACCGCGGTACCCCGCG 1631339
QY 206 CGATCGACGTCGAGGCGCGCGGAGTCTGACGCGCGCTGACCTGCGCGCGTCTGCG 265
Db 1631340 CGGGGCGCGGTTTCCGCGCTGGCAGCGCGTGGCGCGGTGACCGCGGTGCGCG 1631399
QY 266 GACGAGCGTCCACCCCGAGCTCTCTTCTACCGCTGCGCGAACCAGGCGTGGAGCT 325
Db 1631400 TCGTGMGATGCTGCGCGCGCGCGCGCTTCCCTGCGGTGAGCGCGTGGCGCGG 1631459
QY 326 TCGTGGCGCGCGCGTTCACGAGAGAACGAGTATGACGCGCGCGCATCTACGCGCG 385
Db 1631460 CGCGCGGACCGCGGCTTGGCGCGGTGCGCGCTGCGCGCGTGGCGCGCGCGCG 1631519
QY 386 CCGTGTCTCAGGAGCGCGCGCGGAGTCTGCCCTGATCTTCGGCGGTGCTGCGCG 445
Db 1631520 GGAAGGCGCGTCTCTTGGCGCGTGGCGCGGTGGCGCGGTGGCGGAGCGCGCG 1631579
QY 446 CTGAGCGCGGCGATCGCTCGCGGCTCGCTCGCGTCTACCATGACGCGCGCGCA 505
Db 1631580 TCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGACTACGCTTGAACCGCG 1631639
QY 506 GCGGCTGGCTCTGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 565
Db 1631640 CGCGCGCGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1631699
QY 566 ACCTGCACTCCGACGAGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 625
Db 1631700 GCGGACCG 1631756
QY 626 GCGTGGCTCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 685
Db 1631757 AGCGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1631816
QY 686 CCGGGTGGCGGTCTGTCGCGACCGCGAGGAGGTTTCGGCGCGCGCGCGCGCG 745

Db 1631817 CCGGCGCGCAC--CGTTGCGCGCGGGTTACCGTTTGGCGCGCTTTCACGAGGTGGTG 1631873
QY 746 CGCTCGCGACCGCGAGCGCGCGTCCGCAAGTACAGGAGCAGTACTCTACTGGT 805
Db 1631874 GCGTTGGCACTATGCCACCAACGCGCGGTGCGCGCGCGCGCGCGGTGGCGCGCG 1631933
QY 806 CGACCGAGCGCGGTTCACGTTTACCGCGCTGCGCGCGCGCGCGCGCGCGCGCG 865
Db 1631934 CCGGCGCTGCGCGCGTTGCGCGCATTTGCGCGCGTTCGCGCGCGCGCGCGCG 1631993
QY 866 GTGCGCGCGCGACACCGCGCGCGAGATCGAGAACGCGGAGTGGTTCGCGAGACCG 925
Db 1631994 GAGTTGCGCGCGTTGCGCGCGTTCGCGCGCGTTCGCGCGCGCGCGCGCG 1632053
QY 926 TCCACTGGGACACCGCGCGCGGAGATGGGTCTCGCGTATCGGACGACTGCGGTAC 985
Db 1632054 GCGCGCTTGGCGCTGGGTAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632113
QY 986 CGTTCTACACTCGCGAACCTGCGGACTGGGACTTTCG--CGCAACTTCGACTACC 1042
Db 1632114 GGGTTACCGCGCTTACCGCGGTGACCGCGCTTACCATCGCGAAGCGCGGTG 1632173
QY 1043 CGAACACCGCGTTCGCGCGCATCGAGTGGCGCGCGCGCGCGCGCGCGCGCGCG 1102
Db 1632174 GCGCGCTTGGCGCGCTACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632233
QY 1103 GGACACCGCACTGGGTCTGCGCGCGAGCATGGAGCGCTACGCGCGCTCCCGCAT 1162
Db 1632234 ACACCTGATTCGCGTTCGCGCGAAGAGGTTCGCGCGCGCGCGCGCGCGCG 1632293
QY 1163 CGTACGCGTACTGGACAGGACCTGGGACGCGGAGAGTTCACGCGCGCGCGCG 1222
Db 1632294 CCGCGCTTACCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632353
QY 1223 CGCAATGGCTCGACTGGGCTGGGACTGGTACGCGCGCGTACCTGCCATCGACG 1282
Db 1632354 GCGTGGTGGCGTGGCGCGGTGGCGCGGTTCGCGCGCGCGCGCGCGCGCG 1632413
QY 1283 CGCGCGAGACGAGCGCTCGCATCGCTGGATGAACAATGGAAGTACGCGCGCG 1342
Db 1632414 CCG 1632470
QY 1343 ACCTCCCGACGAGTATCGCGCGTACAAAGGCGAGAACTCGATCGTCCGCGAGTGC 1402
Db 1632471 CGCGCGCGCGCGCGCGCGCGCGCGCGTTCGCGCGTTCGCGCGCGCGCG 1632530
QY 1403 GGTGCGCGCGCGCGCGCGCGCGCGCGCGTTCGCGCGCGCGCGCGCGCGCG 1462
Db 1632531 GCGGAGGAAACCGCGCGACCGCGCATGCGCTCCCGCGCTTTCGCGCGCGCG 1632590
QY 1463 CGAACTACGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1522
Db 1632591 CGAAAGCG 1632650
QY 1523 TGCATGGAACG 1582
Db 1632651 GCGTTGCGCGCGTTCGCGCGGTGCGCGCGTTCGCGCGCGTTCGCGCGCGCG 1632710
QY 1583 ACGTGCGCATCTCGTGGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1642
Db 1632711 GCACCGCTGCAACCGCGCGGTGATCGCGCGCGCGCGCGCGCGCGCGCG 1632770
QY 1643 ACGGAGCAGACCTGTAGTGCACCGGAGCGCGCGCGCGCGCGCGCGCGCGCG 1702
Db 1632771 CGGCTACCGCGCGTTCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCG 1632830
QY 1703 CCTACTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1762
Db 1632831 TTGGACCGTTCGCGCGCATACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632890
QY 1763 TCGTGCAGACCGGAGCGGTGAGGTCTTCGTAACGCGCGCGCGCGCGCGCG 1822

Db 926 CACGTCGCGCGCATGGTCGTGACAGCGTGTGTCACACCGTCGCGCGCAAGATCTGGTAC 985
Qy 1510 GGCACGCGCGTCTCCATGGAACGCGCGCATACGAGATCGAGATCGACATCGCCTGG 1569
Db 986 CAGGCAACCTGGACAGAGCGTCCCTTCGAGGCGCGCTGGAAGACTGGAGACTGG 1045
Qy 1570 GACACCGGAGAGAGTGGGATCTCGGTGGGCGCGTCCCGGAGGAGCGGCGCACAG 1629
Db 1046 GTCGCGGAGAGAGCGCGCTTACACCTCGCGGAGACCGCGCGGAGTTCAGGAGCAG 1105
Qy 1630 AACATCGCAAGTACGAGAGAGAGTGTACGTGCGAGCGGAGACCTCGCACTCGCGGG 1689
Db 1106 TGGCTGAGCTGCG 1165
Qy 1690 TACTCGCTGCG 1749
Db 1166 GAGCTGATCTCTCTTCCAGAGCGCGCGCTACTAGGACTCGCGCTGGCGCGCGCGG 1225
Qy 1750 CACCTGCGCATCTCTGTCGACACCGAGCGTGTGAGGTCTGTCACACCGCGCGCACAC 1809
Db 1226 GAGATCTTTCAGCAAGTACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1285
Qy 1810 GTGCTCTCCAGAGGTCCACTTCCCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCG 1869
Db 1286 GACCTGTCGAGACCG 1345
Qy 1870 GCG 1304
Db 1346 GTCGAGTGCACCG 1380

RESULT 14

US-08-387-942C-1
; Sequence 1, Application US/08387942C
; Patent No. 5939289
; GENERAL INFORMATION:
; APPLICANT: ERTESVAG, HELGA
; APPLICANT: VALLA, SVEIN
; APPLICANT: SKJAK-BRAEK, GUDMUND
; APPLICANT: LARSEN, BJORN
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES
; TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: P.O. BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/387,942C
; FILING DATE: 09-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1809-106P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12588 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:
; ORGANISM: Azotobacter vinelandii
; STRAIN: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 290..1951
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2227..6438
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6702..9695
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9973..12588
US-08-387-942C-1

Query Match 4.0%; Score 80.4; DB 2: Length 12588;
Best Local Similarity 42.3%; Pred. No. 1.7e-06;
Matches 508; Conservative 0; Mismatches 691; Indels 3; Gaps 1;

Qy 705 CACCGCGAACACGCGAGGTTTCGGCGCGCGCGTCTGCGCGCTCGCGACACCGACCGAC 764
Db 6875 CATCAGAGCAACGTCATATCTGTCGGCGCGGGGATGGCGAGAGCGTCAATCAAGCTGGT 5934
Qy 765 CGACGGCTTCGCGCAAGTACCAGGACGATCTTACTTGTGTGACCGGCGGGTTCAC 824
Db 6935 CGACGGCTGGGATCAGGACGTCACCGGCATCGTCCGCTCGGCTACGGCGAGGAGACCG 6994
Qy 825 GTTCACCGCGCTCGCGACCGCGCTCATGTCACACCGACCGGTGCGCGCGCGACCGCC 884
Db 6995 CAACCTTCGCGATGAGCGACCTGACCTCGCGCAACCGCGCAACACGCGCGCAAGGT 7054
Qy 885 CGCGGAGATCGAGAACGCGGAGTGTTCGGCGACCGCGCGCGCGCGCGCGCGCGCGCG 944
Db 7055 CGACGGCTGTTCAACGGCTACATTCGCGCGGAGGACCGCGCGCGCGCGCGCGCGCG 7114
Qy 945 CG---GAGAAATGGGTCTCGGTCTCGGACGACTGCGGTGACGCGCGGTTCCTACACCTCGCC 1001
Db 7115 GGACGGGTGGAAATCCGTTGAAATGTCGGTTTACGTTTCGATCGCGCGCGCGCGCGCG 7174
Qy 1002 GAACCTGCGGACGTCGACACTTCGCGCAACTTCGACACTACCGCAACCGCGCGCGCG 1061
Db 7175 CAACCTGACGATCGCGACGAGGTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7234
Qy 1062 CATCGAGTCCCCCGACCTGTTTCGAGATCACCGGACGAGCGGACCGCGCGCGCGCGCG 1121
Db 7235 TTTCAGATCGCGGGGTGTTTCGAGAACACGCTCTCGTACAAACACGCGCGCGCGCGCTT 7294
Qy 1122 CGCGCGCGATGAGCGCTTACGGGATCGGCTTCCCGATGAGGTACGCTTACTTGACAGG 1181
Db 7295 CAACATCGTCAACGACCAACGACTTCTGCTTGAGCAACACGCTGCGCTTTCGCGAAGCG 7354
Qy 1182 CACTGGGACGCGGAGCAGTTCACGCGCGAGCGCTTCCCGCGCGCGCGCGCGCGCGCGCG 1241
Db 7355 CGGCGCGCGCGCTGTTGTCAGCGCGGCTCGTCCGACGTGGCGCGCGCGCGCGCGCGCT 7414
Qy 1242 CTGGGACTGTACGCGCGCGCTCACCTGGCCATCGATCGACGCGCGCGCGCGCGCGCGCT 1301
Db 7415 GATCGAGCGCGCGCTTACTAGGACACGCGCTGGAAGCGGTGACAGATCAAGATGCGCCA 7474
Qy 1302 CGCATCGCGGTGATGAACAACTGGAAGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1361
Db 7475 CGAGTCAACCTCGAGAACGCGGAGATCTACGGCAACCGCGCTATACGGGGTTCGCGGTCTA 7534
Qy 1362 CGACGGCTACAAACGGGAGAACTCGATCGTCCGCGAGCTGCGGCTCGCGCGCGCGCGCG 1421
Db 7535 CGGCGCGGAGGATGTGAGATCTTCGCAACTATACATTCACGCAATTCGCGAGACGGTTC 7594
Qy 1422 CGGTGTGTACACCTCTCTGAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1481
Db 7595 CTACGCGGAATCTCTCTGAGTCTTACGAGATACCGCGCGCGCGCGCGCGCGCGCGCG 7654

Qy	1482	CACCACACTCCCGACGGGACCGTGCAGCGCAGCGCGTCTCTGCCATCGAAGACGGACGGCG	15414
Db	7655	CACCACACCGGACCGTGGATCGAAGGCCAACCACTCGTCGGCTCGGCCAACTCCACCTA	7714
Qy	1542	ATACGAGATCGAGCTCGACATCGCTGGGACACCGCGAGAACGTCGCGCATCTCGGTGGG	1601
Db	7715	TGGATCCAGGACGGACGAGCGGCACCGACTTACAGCAGCCTCTACGCCCAACAGCGTCAG	7774
Qy	1602	CCGCTCCCGACGACACCCGGCACACGNAACATCGCGAAGTACGAGACGACACTGTACGT	1661
Db	7775	CAATGTGCAGAACGGCTCGGTGGCTCTACGGCGCCAACTCCGTCGCTCCCGACCTGCC	7834
Qy	1662	CGACCGAGGACCTCCGACCTCGCGGGGTACTCGTGTGCGCCCTACTCTCGCAGCGCGCGC	1721
Db	7835	CGGACCGGGCAGCAGCGGACCTCGAAGGCACGCGCCGGCACACGACAGCTTTGGCGCAG	7894
Qy	1722	CCGCATCGACCCCGGCGCCGATCCGTGCACCTCGCGCATCTCTGTCGACACCCAGAGCGT	1781
Db	7895	CGACGCCAGGAGACGCTGCTGGGCTGGACGGCAACACCGCTGACCGCGCGCGCGG	7954
Qy	1782	CGAGGTCTTGTCAAGCGCGGCCACACCGTGTCTCCAGCAGAGTCCACTTCGCGCGAGGG	1841
Db	7955	CAACGACATCTCTGACGGGGCGCGGGCGGCACAACTGACCGGCGCGCGCGCGCGCA	8014
Qy	1842	CGACACGGGAATCTCTCTACACGACGCGGGCGCCGACACTTACCGGCATCTGCTCGT	1901
Db	8015	CTGTCTCCGGTCTCCGCGGCACCGCAGCTACCGCACCGCAGCGCCAGCTTCAACGA	8074
Qy	1902	CC 1903	
Db	8075	CC 8076	

RESULT 15

US-09-105-537-1

Sequence 1, Application US/09105537A

: Patent No. 6265202

; GENERAL INFORMATION:

APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

APPLICANT: Zhao, L.

TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.438US1

; CURRENT APPLICATION NUMBER: US/09/105,537A

; CURRENT FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSEO for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 15872

GENOTYP: IS
TYPE: DNA

ORGANISM: Streptomyces venezuelae

US-09-105-537-1

Query Match	4.0%	Score 80.4;	DB 4;	Length 15872;
Best Local Similarity	42.7%	Pred. No. 1.7e-06;		
Matches 649;	Conservative	0;	Mismatches 856;	Indels 15;
Gaps				4;

Qy	492	CATGACGCCCCCAGCGGTGCTCTGGACACCCCAACGCCCGGTACACCAACCGGCGC	551
Db	472	CATCGGCACCGCGGTCTCGTATCACCTCGGCCTGCAAGGCCCGAGCTCACGTCGACGC	531
Qy	552	CTACAGCTCTACTACTGTCACCTCCGACACAGAACAGGCCCGCGGCTGGGACCAACGC	611
Db	532	CGGCGAGTGTCTCGGTCTGCTCCCGCTGCACCTGGCCTTGGAGTTCCTTCGCGCGCCGGGA	591
Qy	612	GAGCAGCAGCGACGGGTGCCTTTCACGCACACGGCACCGCTGATGCCGCTCGGCGCGGA	671
Db	592	GTCCAGCAGCGGCGTCTGTCGCGGCGTGAACCTCAACATCCTCGGGAGAGCCCGGTGAC	651
Qy	672	CTTCCCGGTGTGTCGGGTTCGGGTCTGTGGCACCGGACACAGGCAGAGGTTTCGGGCGC	731
Db	652	GGAGGAGCGGTTCGGTGGACTCTCCCGGACGAGCACCGCTACACCTTCGAGCCGCGGGC	711

Result No.	Score	Query %		DB	ID	Description
		Match	Length			
1	97.6	4.9	2712	10	US-09-748-033-4	Sequence 4, Appli
2	88.2	4.4	882	10	US-09-074-300-735	Sequence 735, App
3	80.4	4.0	15872	9	US-09-860-846-1	Sequence 1, Appli
4	80.4	4.0	15872	9	US-09-988-384B-1	Sequence 1, Appli
5	80.4	4.0	15872	10	US-09-861-289-1	Sequence 1, Appli
6	80	4.0	1896	9	US-10-124-880-15	Sequence 15, Appli
7	79.2	4.0	2334	10	US-09-476-242-7	Sequence 7, Appli
8	78.2	3.9	11220	9	US-09-860-846-32	Sequence 32, Appli
9	78.2	3.9	11220	9	US-09-988-384B-32	Sequence 32, Appli
10	78.2	3.9	11220	10	US-09-861-289-32	Sequence 32, Appli
11	78.2	3.9	36778	9	US-09-860-846-5	Sequence 5, Appli
12	78.2	3.9	36778	10	US-09-861-289-5	Sequence 5, Appli
13	78.2	3.9	37948	9	US-09-988-384B-5	Sequence 5, Appli
C 14	77.6	3.9	12441	9	US-09-988-384B-3	Sequence 3, Appli
C 15	77.6	3.9	13613	9	US-09-860-846-3	Sequence 3, Appli
C 16	77.6	3.9	13613	10	US-09-861-289-3	Sequence 3, Appli
17	75.8	3.8	2322	10	US-09-476-242-18	Sequence 18, Appli
18	75.8	3.8	2322	10	US-09-476-242-19	Sequence 19, Appli
19	75.6	3.8	4689	9	US-09-860-846-34	Sequence 34, Appli

	Query Match	4.9%;	Score 97.6;	DB 10;	Length 2712;
	Best Local Similarity	43.9%;	Pred. No. 1.6e-13;		
	Matches 650;	Conservative 0;	Mismatches 794;	Indels 36;	Gaps 4;
QY	466	CGGGCGTCGTCGGTCCGCTACACATGACGCCGCCACGGGTGGCTCTCGACACCC	525		
Db	845	CGGGCGCGCTGGCGCGGCTGTCGGCTCGCGTCCCTTCGGCGGGATGGTCGGCCTC	904		
QY	526	CACAGCCCGGTACACACCCACGGCGCCTACAGGTGTACTACTGTACCTCGACACAGAAC	585		
Db	905	CGCGCCCCACCCAGGCGCGCGAGCGCGCGCCAGCACCCACGGCTCCTACACAGGACG	964		
QY	586	AACGGCCCCGGCGCTGGGACACGCGAGCAGCACGACGGCGTGCCTTTCACGGACCCAC	645		
Db	965	CAGGACTGGGCGACCGGCTTCAGGGCAATGGAGCGTGAAGACACCGGCACCCGCCCC	1024		
QY	646	GGCACCGTGATGCGGCTCGCGGCCGACTTCCCCGTGTGGTCCGGGTGCGCGGTGCTCGGC	705		
Db	1025	CTCAGCGGCTGGACCCCTGGAGTGGGACTTCCCCCGGGAACCAAGGTGACCTCGCGCTGG	1084		

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 1

LENGTH: 15872

TYPE: DNA

ORGANISM: Streptomyces venezuelae

US-09-860-846-1

Query Match 4.0%; Score 80.4; DB 9; Length 15872;
Best Local Similarity 42.7%; Pred. No. 1e-09;
Matches 649; Conservative 0; Mismatches 856; Indels 15; Gaps 4;

QY 492 CATGACGCCCCAGCGCTGGCTCTGCGGACCCCAAGCGCCGCTACCAACCCAGGCGG 551
DB 472 CATGCGCAACCGCTCTGTAACCACTCGGCTCGAGGGCCCGAGCCTCACCGTCGACGC 531
QY 552 CTACGAGCTGTAACCTGCACTCTGCGACCAAGAAAGCGCCCGCGCTGGGACCAAGC 611
DB 532 CGCGAGTCGTCCTGCTGCTGCGCGTGCACCTGGCTGCGAGTCCCTGCGCGCGGGA 591
QY 612 GAGCAGACCGAGCGGTGCGCTTTCAGCACCAGCGACCGTGTATGCGGTGCGGCGCGA 671
DB 592 GTCCACGACGCGCTGCTGCGCGCGGTGAACCTCAACATCTCGGAGAGCGCGGTGAC 651
QY 672 CTTCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 731
DB 652 GGAGAGCGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 711
QY 732 CGGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 791
DB 712 CAACGATTGCTGCGGCGGAGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 771
QY 792 GTACTCTACTGTGCGACGCGCGGTTCACGTTTCAACCGCTGCGCGCGCGCGCTCAT 851
DB 772 CTTCCGCGACGCGCGGTGCTGCGCGCGGTGCTGCGCGCGGTGCTGCGCGCGGTGCT 831
QY 852 CGTCAACACCGAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 908
DB 832 AGCCACCGCGGTGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 891
QY 909 GTTCCGCGACCGCAAGATCCACTGCGGACACCGCGCGCGGAGATGGTCTGCGTCACTCG 968
DB 892 GTACCGGAGCGCGCTGAGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCG 951
QY 969 ACGACTGCGGTACGCGCGGTTCATACCTGCGCGAACTGCGCGAGCTGGACACTTCGCG 1028
DB 952 AACCCCGCTCGCGACCGCATCGAGCGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCG 1011
QY 1029 CAACTTCTGACTACCGAACCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGAG 1088
DB 1012 CCGCGCGAGCAACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1071
QY 1089 CACCGCAG--ACGACGGGACACGCCACTGGGTGCTGCGCGCGCGAGCTGAGCGCTACGG 1145
DB 1072 CGCGCGCGCATCTGCGGCTCATCAGACGCTCTGCGGTGCGCGCGCGCGCGGATGCC 1131
QY 1146 CATCGGCTCTCCCATGAGCTACGCTTCTGAGACGACCTGGGACGCGGCGAGCTTCCA 1205
DB 1132 GCGAGCTCAACTTCGCTAGCGCGCGCGCGGACATCCGCTCGACACCTCTGCGGCTCGA 1191
QY 1206 CGCGCAGCACTTACCGCGCAATGGCTGCTGAGTGGGTGCGGTGCTGAGTGGCGCGCTCAC 1265
DB 1192 CGTCCCGGACGCGCTGCGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGAG 1251
QY 1266 CTGCGCATGATCAGCGCGCGGAGACCAAGCGCTGCGCATCGCTGCGGTGAGTGAACACTG 1325
DB 1252 CTCGTTGCGGATGGCGGCGCAACAGCGCGCGCGCGCTGCTCAGCAAGCGCGCGCGCG 1311
QY 1326 GAAGTACGCGCGACGCGACGCTCCCGACCGCGCATCGACGCGGTACACGGCGGAGAACTC 1385
DB 1312 CGGCGAGCAGCGCGGATCGATGAGGAGACCGCGCTGAGACGCGGCGCGCGCGCTT 1371
QY 1386 GATCGTCCGCGAGCTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1445

DB 1372 CGTCTCACCGGCG 1428
QY 1446 CCCGCTGGCAGCGCTGAGCACTACGTACCGCGCACACCACTACCTCCCGCGACGACGCT 1505
DB 1429 GCGCTGCAAGCGGACCGCGGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1488
QY 1506 CGAGCGGAGCGCGCTGCTGCGCATGGAACGAGCGCGCGCGCATGAGATCGAGTTCGATCGC 1565
DB 1489 CCGTACGCTCTTACGACCGACCGGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 1548
QY 1566 CTGGGACACCGCGAGCGATCGGTCGCGCATCTGCTGCGCGCGCGCGCGCGCGCGCGCG 1625
DB 1549 CGGCTCGGCGCGCTGCG 1608
QY 1626 CACGAACATCGCAAGTACGAGCAGACCTGTACGTGCGACGAGGACCGCTCCGACCTCGC 1685
DB 1609 CCGCGGCGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1668
QY 1686 CGGTACTGCTGCGCGCGCTTACTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1745
DB 1669 GGAGTTGCTACG 1728
QY 1746 GGTCACTGCGCATCTGCTGCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1805
DB 1729 GGACCGCGCTGCGACCG 1788
QY 1806 CACGCTGCTCTCCCGAGCGAGTCCACTTTCGCGGAGCGCGCGCGCGCGCGCGCGCG 1860
DB 1789 CACGCTCCACACAGCGCGCGCGCTTTCGCGGTGAGGTGCGCGCTCCACCGCTCGT 1848
QY 1861 -TACACCGAGCG 1919
DB 1849 GTCTGCGGCGCTACG 1908
QY 1920 GATCTAGCGATGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1979
DB 1909 CCACGTGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1968
QY 1980 CAATCGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1999
DB 1969 CCTCATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT

RESULT 4

US-09-988-384B-1

Sequence 1, Application US/09988384B

Publication No. US20030073824A1

GENERAL INFORMATION:

APPLICANT: Sherman, D.H.

APPLICANT: Liu, H.

APPLICANT: Xue, Y.

APPLICANT: Zhao, L.

TITLE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 600.536US1

CURRENT APPLICATION NUMBER: US/09/988, 384B

CURRENT FILING DATE: 2001-11-19

PRIOR APPLICATION NUMBER: PCT/US99/14398

PRIOR FILING DATE: 1999-06-25

PRIOR APPLICATION NUMBER: US 09/105,537

PRIOR FILING DATE: 1998-06-26

NUMBER OF SEQ ID NOS: 53

SEQ ID NO 1

LENGTH: 15872

TYPE: DNA

ORGANISM: Streptomyces venezuelae

US-09-988-384B-1

Query Match

Best Local Similarity 4.0%; Score 80.4; DB 9; Length 15872;

Matches 649; Conservative 0; Mismatches 856; Indels 15; Gaps 4;

QY 492 CATGACGCCCCAGCGCTGGCTCTGCGGACCCCAAGCGCCGCTACCAACCCAGGCGG 551
DB 472 CATGCGCAACCGCTCTGTAACCACTCGGCTCGAGGGCCCGAGCCTCACCGTCGACGC 531
QY 552 CTACGAGCTGTAACCTGCACTCTGCGACCAAGAAAGCGCCCGCGCTGGGACCAAGC 611
DB 532 CGCGAGTCGTCCTGCTGCTGCGCGTGCACCTGGCTGCGAGTCCCTGCGCGCGGGA 591
QY 612 GAGCAGACCGAGCGGTGCGCTTTCAGCACCAGCGACCGTGTATGCGGTGCGGCGCGA 671
DB 592 GTCCACGACGCGCTGCTGCGCGCGGTGAACCTCAACATCTCGGAGAGCGCGGTGAC 651
QY 672 CTTCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 731
DB 652 GGAGAGCGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 711
QY 732 CGGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 791
DB 712 CAACGATTGCTGCGGCGGAGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 771
QY 792 GTACTCTACTGTGCGACGCGCGGTTCACGTTTCAACCGCTGCGCGCGCGCGCTCAT 851
DB 772 CTTCCGCGACGCGCGGTGCTGCGCGCGGTGCTGCGCGCGGTGCTGCGCGCGGTGCT 831
QY 852 CGTCAACACCGAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 908
DB 832 AGCCACCGCGGTGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 891
QY 909 GTTCCGCGACCGCAAGATCCACTGCGGACACCGCGCGCGGAGATGGTCTGCGTCACTCG 968
DB 892 GTACCGGAGCGCGCTGAGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCG 951
QY 969 ACGACTGCGGTACGCGCGGTTCATACCTGCGCGAACTGCGCGAGCTGGACACTTCGCG 1028
DB 952 AACCCCGCTCGCGACCGCATCGAGCGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCG 1011
QY 1029 CAACTTCTGACTACCGAACCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGAG 1088
DB 1012 CCGCGCGAGCAACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1071
QY 1089 CACCGCAG--ACGACGGGACACGCCACTGGGTGCTGCGCGCGCGAGCTGAGCGCTACGG 1145
DB 1072 CGCGCGCGCATCTGCGGCTCATCAGACGCTCTGCGGTGCGCGCGCGCGCGGATGCC 1131
QY 1146 CATCGGCTCTCCCATGAGCTACGCTTCTGAGACGACCTGGGACGCGGCGAGCTTCCA 1205
DB 1132 GCGAGCTCAACTTCGCTAGCGCGCGCGCGGACATCCGCTCGACACCTCTGCGGCTCGA 1191
QY 1206 CGCGCAGCACTTACCGCGCAATGGCTGCTGAGTGGGTGCGGTGCTGAGTGGCGCGCTCAC 1265
DB 1192 CGTCCCGGACGCGCTGCGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGAG 1251
QY 1266 CTGCGCATGATCAGCGCGCGGAGACCAAGCGCTGCGCATCGCTGCGGTGAGTGAACACTG 1325
DB 1252 CTCGTTGCGGATGGCGGCGCAACAGCGCGCGCGCGCTGCTCAGCAAGCGCGCGCGCG 1311
QY 1326 GAAGTACGCGCGACGCGACGCTCCCGACCGCGCATCGACGCGGTACACGGCGGAGAACTC 1385
DB 1312 CGGCGAGCAGCGCGGATCGATGAGGAGACCGCGCTGAGACGCGGCGCGCGCGCTT 1371
QY 1386 GATCGTCCGCGAGCTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1445

Db 472 CATCGCCAAACCCGGTCTCGTACCACTCGGCTCGAGGCGCGGACCTCACCGTCGAGCG 531
QY 552 CTACAGCTGTACTACCTCGACTCGACAGAACAGCGCCCGCGGCTGGGACAGCG 611
Db 532 CGCGAGTGTCTCTCGTCTGCGCGTGCACCTGGCTCGGAGTCCCTGCGCGCGGGA 591
QY 612 GAGCAGGACCGGCTGCTGCTTACGAGCAGCGGACCGTGTGCTGCGGCGCGA 671
Db 592 GTCCAGAGCGGCTGCTGCGCGGCTGAACCTCAACATCTTCGCGGAGAGCGCGTAC 651
QY 672 CTTCGCCGTGTCTCGGCTCGGCTCGTCCGACCGGACAGCGGAGTTCGCGCG 731
Db 652 GAGAGAGCGCTTCGCTGAGTCTCCCGAGCGGACCGCTACACTTCGAGCGCGGCG 711
QY 732 CGCGCGGCTCGTGGGCTCGGACCGACCGGACCGGCTCGGACCGGCTCGGACGAGCA 791
Db 712 CAAGGATTCGTCGCGGCGGAGCGGCGGAGTCTGCTCAAGCGCTCTCCCGCGC 771
QY 792 GTACCTTACTGCTGACGAGCGGCTTACGTTTACCGGCGCTTCCCGACCGCTCAT 851
Db 772 CTTCCCGACGCGGCTGTCTGCGGCTCATCGGCTCATCGGCGCGGCTCAACAGCGG 831
QY 852 COTCAACACGAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 908
Db 832 AGCCACCGGCTCTACCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 891
QY 909 GTTCGCGGACCGGCTTCTACCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1028
Db 892 GTACCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 951
QY 969 AGGACTGCGGTACGCGGCGGCTTCTACCTCGGCGGCGGCGGCGGCGGCGGCGG 1028
Db 952 AACCGCGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1011
QY 1029 CAATCTGACTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1088
Db 1012 CCGCGGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1071
QY 1089 CACCGCAG---ACGAGGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1145
Db 1072 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1131
QY 1146 CATCGGCGCTCCCATGAGTACGCTTCTGAGGAGCGGCGGCGGCGGCGGCGG 1205
Db 1132 GCGAGCGCTCACTTCGCTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1191
QY 1206 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1265
Db 1192 CGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1251
QY 1266 GTGGCGATCGATCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1325
Db 1252 CTCGTTGCGGATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1311
QY 1326 GAGTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1385
Db 1312 CGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1371
QY 1386 GATCTGTCGCGGAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1445
Db 1372 CGTCTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1428
QY 1446 CCGCGTGGAGCGGCTGAGGAGTACGTCAGCGGCGGCGGCGGCGGCGGCGG 1505
Db 1429 GCGCGTGAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1488
QY 1506 CGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1565
Db 1489 CGGTACGCTTTCAGCGACCGGCTGCTGCTCGCGGCGGCGGCGGCGGCGG 1548
QY 1566 GTGGGACACCGCGGAGCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1625
Db 1549 CGGCGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1608

QY 1626 CACGAACATCGCAAGTACGGAGCAGACCTGTACGTGACCGAGGACCGCTCCGACCTCGC 1685
Db 1609 CCGCGGCGGCGGCTCGCGCTCTTTCAGCGGCGGCGGCGGCGGCGGCGGCGG 1668
QY 1686 CCGGCTACTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1745
Db 1669 GAGGTTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1728
QY 1746 CCGTACCTGCGCATCTCTGTCGACACAGCGGCTCGAGGCTTCTGTCGTCGCGGCGG 1805
Db 1729 GAGCGGCTCTCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1788
QY 1806 CACCGTGTCTCTGACGAGGCTCTGCGGCGGCGGCGGCGGCGGCGGCGG 1860
Db 1789 CACCGTCTACACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1848
QY 1861 -TACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1919
Db 1849 GTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1908
QY 1920 GATCTAGGCGATGACACACAGCGGCTCACGAGCGGCGGCGGCGGCGGCGG 1979
Db 1909 CCAGTCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1968
QY 1980 CAATCGACAGCTCTCTGTCG 1999
Db 1969 CCTCATGCGGCGGCTCCCGG 1988

RESULT 5
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1

Query Match 4.0%; Score 80.4; DB 10; Length 15872;
Best Local Similarity 42.7%; Pred. No. 1e-09; Indels 15; Gaps 4;
Matches 649; Conservative 0; Mismatches 856;

QY 492 CATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 551
Db 472 CATGCGCAACCGGCTCTGTACCACTCGGCTGCGGCGGCGGCGGCGGCGG 531
QY 552 CTACAGCTGTACTACTGTGACTCGGACGAGAACAGCGGCGGCGGCGGCGG 611
Db 532 CCGGAGCGTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 591
QY 612 GAGCAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 671
Db 592 GTCCAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 651
QY 672 CTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 731
Db 652 GAGGAGCGCTTCGCTGGACTCTCCCGGAGCGGCGGCGGCGGCGGCGGCGG 711

QY 732 CGGCGCGGTCTCGCGTTCGGAGCCAGCCAGCCAGCGGGGTTCGCAAGTACACGAGGACA 791
Db 712 CAACGGATTCGTCGGGCGAGGCGCGGAGTCTGCTACTCAAGCGGTCTCTCCGCGC 771
QY 792 GTACCTCTACTGTTGACGCGGCGGTTTACAGTTTACCGCCCTCCCGACCCCGTCAT 851
Db 772 CTTCCGACGCGGACCGTGTCCAGCGGTATCCGCGCCAGCGCCGTCAACAACGACGG 831
QY 852 CGTCAACACCGAGCGTTCGGCGCCAGCCAGCCCGCC---GAGATCGAGAACGCCAGTG 908
Db 832 AGCCACCCGGGTCTACCGGTGCGCCAGCAGGCGCCCGCAGAGAGGTGCTGCGCGAGC 891
QY 909 GTTCCGCGACCCCAAGATTCACCTGGGACACCGCCCGCGGAGAAATGGTCTGCGTATCGG 968
Db 892 GTACCGGAAGGCGGCGTGGACCGGTGCGCGTCCAGTACGTCCAGAACTCCACGGCACCGG 951
QY 969 ACGACTGGGTACCGCGGTCTACACTCGCCGAACCTGCGCGACTGGACACTTTCGCGC 1028
Db 952 AACCCCGTGGCGACCCATCGAGGCGCGCGCTCGGCGCCGTCTCGGCTCGGCGCG 1011
QY 1029 CAACTTCGACTACCCGGAACACCGCCCTCGCGGCATCGAGTGCCTCCGACCTGTTTCGAGAT 1088
Db 1012 CCCGCGGACGAAACCCCTGCTGCTGCTCGGCCAAGACGACGTCGGGCACTCGAAGG 1071
QY 1089 CACCGCAG---ACGACGGGACACGCCACTGCGGTGCTCGCGCCAGCATGACGCTTACGG 1145
Db 1072 CGCGCGCGGATCTGCGCGCTCATCAAGACGCTCTCGCGCTCGCGCGCGCGGATCCC 1131
QY 1146 CATCGGCTCCCATGACGTACGCTACTGGACAGGACCTGGACGCGGAGCATTTCCA 1205
Db 1132 GCGAGCGCTCAACTTCGCTAGCGCCCAACCGGAGCATCCGCTCGACACCTTCGGGCTCGA 1191
QY 1206 CGCGGAGGACCTACCCCGCAATGGCTCGACTGCGGGTGGGACTGTACGCGCGCTCAC 1265
Db 1192 CGTCCCGACGGCTGCGGGAGTGGCGCACCGGACCGGAACTCTCGCGCGCTCAG 1251
QY 1266 CTGCGCATGATCAACGCGGCGGAGACAGAGCGCTCGCATCGCGTGGATGAACAATG 1325
Db 1252 CTCGTTGCGGATGGCGGACCAACGCGCCAGCTGCTCTAGCGAAGGCGCGCCCGAGG 1311
QY 1326 GAAGTACGCGCAGCGACGCTCCCGACGAGCATCGCGGCTACACGGGAGAACTC 1385
Db 1312 CGGCGAGCGCGCGGATCGATGAGGAGACCCCGCTCGACAGCGGGCGCGCACTGCGCTT 1371
QY 1386 GATCGTCGGCGAGTTCGGGTCGCGCCGACAGCGCTGGGCGGTGTACACCTCTCTGAGCAC 1445
Db 1372 CGTGTACCGCGCGCGCGGAGGCGCTTCG---GGCGCCAGCGCGCGGCTGACGA 1428
QY 1446 CCGCGTGGCAGCGGTGACGAATACGTACCGGACACCACTCCCGACCGGACCGT 1505
Db 1429 GCGCGTCGAAGCGGACCGGAGCTCGCGCGCGCGGCACTCGCGCGGTGCTGTCACAC 1488
QY 1506 CGAGCGGAGCGGCTCTGCTGATGGAACGAGCGGCGCATACGATCGAGTCTGACATCGC 1565
Db 1489 CCGTACGGTCTTACGACCGCGGTGCTGCTCGCGCGGAGCGCGCGCGCTCTCTCGA 1548
QY 1566 CTGGGACACCGGACGAGCTCGGCACTCTCGGTGGGCGGTTCGCCCGGACGAAACCGGCA 1625
Db 1549 CGGCTCGGCGCGCTCGCGCGCGGAGCGCGCGCGCGGCGGTGTACCGGACCGCGCGC 1608
QY 1626 CACGAACATCGGCAAGTACGAGGACGAGACTGTAGCTGACCGGAGGACCGCTCCGACCTCGC 1685
Db 1609 CCGCGGCGGCTCGCGCTCTGTTTACGCGGCGGAGGTGCGCAAGCTACGGGCGATGGGAT 1668
QY 1686 CGGTACTCTCGCGCGCTTACTCGGAGCGCGCGCGCGCGCGCATCGACCGCGCGCGCGATC 1745
Db 1669 GGAGTTGACGCGCGCGCGCGCTTTCGGGAGCGCGCTTTCGACCGCGCTCGCGCGCGAACT 1728
QY 1746 CGTGCACCTGCGATCTCTGTCGACACCCAGAGCGTCTGAGGTCTCTGTCGAACGCGCGGCA 1805
Db 1729 GGACCCCTCTCTGACCGCGCGCGCTCGCGGAACTGCTCGCGGCGGCGGACACCTTCGACCG 1788
QY 1806 CACCGTCTCTCCAGCAGGTTCACCTTCGCGGAGGCGGACACGGGAATCTCGCTC----- 1860

Db 1789 CACGCTCCACACACAGCCGCGCTCTTCGCGGTGAGGTGCGCTCCACCGCTCTGTCGA 1848
QY 1861 -TACACGAGCGGCGCGCCGACACTTCACCGCATCGTGTGCGGAGATTTGGCCAGGC 1919
Db 1849 GTCTGCGGGGTACAGCCGACCTGTCGCGGCGCACTCCGTCGCGGAGATCAGCGCGC 1908
QY 1920 GATCTAGCGATGACACACACCGCTACCGAAAGCGCGCGCCCGGAGAGACGAGCGCGA 1979
Db 1909 CCAGTCCGCGGGTCTGTCGCTGCGGAGCGCGCGCGCTCTGTCGCGCGCGCGCGC 1968
QY 1980 CAATCGACAGTCTCTGTCG 1999
Db 1969 CTTATGACGCGGTCTCCCG 1988
RESULT 6
US-10-124-880-15
; Sequence 15, Application US/10124880
; Publication No. US20030026810A1
; GENERAL INFORMATION:
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene No. US20030026810A1boe
; APPLICANT: Schulein, Martin
; APPLICANT: Outtrup, Helle
; TITLE OF INVENTION: No. US20030026810A1el Rhamnogalacturonan Hydrolases
; FILE REFERENCE: 5572-204-US
; CURRENT APPLICATION NUMBER: US/10/124,880
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US/09/311,626B
; PRIOR FILING DATE: 1999-05-13
; PRIOR APPLICATION NUMBER: 0608/98
; PRIOR FILING DATE: 1998-05-01
; PRIOR APPLICATION NUMBER: 60/084,358
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1896
; TYPE: DNA
; ORGANISM: Streptomyces coelicolor
US-10-124-880-15
Query Match 4.0%; Score 80; DB 9; Length 1896;
Best Local Similarity 43.8%; Pred. No. 1.6e-09;
Matches 665; Conservative 0; Mismatches 825; Indels 30; Gaps 6;
QY 489 CCACATGACGCGCGCGCGCTGCTGCGACCGCCCAAGCGCGCGGTCTACACCCACGG 548
Db 276 CAACTCCAGCGCGCTCACCGGCTCCACACCTACTTCCACTCCGCGCGCGCTCCACGC 335
QY 549 GCGCTACAGCTGTACTACCTGCACTCGGACCGAACAAGCGCGCGCGCGGTGCGACCA 608
Db 336 CGACTACAGCTCGCGCGGTGTAAGCGGACGAGGCGGCGGACTCCCGTCCACGGAT 395
QY 609 CGCGAGCAGCAGCAGCGCGGTTCAGCAGCAGCGGACCGCTGATCGCGTCCGCGC 668
Db 396 CCAGTTCGCGCGCGGTACAGGAGCTACCGATCAGCGCGCGCTCCCGGCGCACCC 455
QY 669 CGACTTCCCGGTGTGTCGCGGTGCGGCGGTGTCGGGACCGCGCAACAGCGGAGGTTTCG 728
Db 456 CGAGCGGTCTCTTACACTACGAGGCGCAACGAGCGCTCGGCGGACCTCGAGCGCA 515
QY 729 CGCGGCGGTGCTGCTCGGCTCGCGCGCGCGCGCGCGCGGTTCGCGAGTACGACCA 788
Db 516 CGCGCGCTCGACCTCTGCTCTCAAGTGGACCGCGCAACGCGGAGCAACTCCCGT 575
QY 789 GCAGTACTCTACTGTCGACCGGCGGTTCACGTTTACCGCTTACCGCGCTGCGCGACCCGT 848
Db 576 CGGCTACCGGCAACAGGTCGTCGAGCGGATCAAGCTCGACGGCA---CCGCGCTGTC 632
QY 849 CATGCTCAACACCGGCTGTCGCGCGCGCGCGCGCGCGGATCGAGAACGCGCGAGT 908

Db 633 GCGGTGACCTGGCGCGAATCCGCTCCGGCGCCACATACACCCAGTTCACAGGTGA 692
QY 909 GTTCGGCGACCCAAAGATCCACTGGGACACCGCCCGGAGAAATGGGTCTCGGTTCATCGG 968
Db 693 CGACTACGACGCGAGCGCGCGCGGAGGTGCCATGAGACCGCGGACCGAACGA 752
QY 969 AGGACTCGGGTACGCGGGTTCTACACCTCGCCGAACCTGGCGGACTGGACATTCGCGG 1028
Db 753 CGGCACCGCGCGGTCTATCGGCAACTCTCGCGGATCACCAGCACTCGAGCGGTACGT 812
QY 1029 CAACTTCGACTACCCGAACACCGCCCTCGGGGCATCGAGTCCCGGACCTGTTTCAGAT 1088
Db 813 CCTCTCCGGCGCGAATACCTACCATGTTCAACGCGCGGACCGGACCGCATGGGAC 872
QY 1089 CACCCAGACGACGCGACACGCACTGGTGTCTGCC---GCCAGCATGGAGCCTACGG 1145
Db 873 CGTCGACTACTCCGCGCGCGGCTCGGTCTCTCTGGGCGACTCTTAGGGCAACG 932
QY 1146 CATCGGCTCCCATAGCTAGCTACTTGGACAGCACTGGGAGCGGAGCGAGCTTCCA 1205
Db 933 COTGACCGGTTCTTGGGGGACGCGGTACCTGGACGCTCCCGCCCTCCGTGATTAT 992
QY 1206 CGCGACGACCTACCCGCAATGCTCGACTGGGCTGGGACTGTACGGCGCTCAC 1265
Db 993 GCGCGGGGTACTACGCGCACCGGTGATCGCGCCTGGGACTGG---CGGGACGGCG 1049
QY 1266 CTGGGCATGATCGAGCGCGCGGAGACCAAGCGCTCGCCATCGGTGTGATGAACAACTG 1325
Db 1050 GTTACCGCGCGCTGACTTTCGACACCAACTCTCCACCAACACGCGCAAGGCTACGA 1109
QY 1326 GAAGTACGCGGACCGGAGTCCCGACCGCATCGAGCGGTACAGCGGCGAGAACTC 1385
Db 1110 CGCGAGGCGCAACCACTGCTCGCTCGCGGACGTGGACGCTGACGGCGGAGAGAT 1169
QY 1386 GATGTCGCGGAGTGGGCTGCGCGACAGCTTGGCGGCTGGTACACCTCCTGAGCAC 1445
Db 1170 GGTCTACGCGCGGATGGCGGTGCGAGCAACGCTTACGCGCTGTGGACCAACAGGAAG 1229
QY 1446 CCCCGTGGCAGCGTACGAACTAGCTCACCGCCACCAACC-----ACACTCCCCGACCG 1499
Db 1230 CCAGCGGACGCCATGCAGCTCGGCGACCTCGACCCGTCGCGGGCGGCTCGAGAGTT 1289
QY 1500 GACCTGACGCGCAGCGCTCTCCATGAGACGCGCATACGAGATCGAGCTGA 1559
Db 1290 CAAAGTTCAGCAGGAGCGCTCGAAGCCCTCGTCTGTAACGCGGACCGCCGACGGGCA 1349
QY 1560 CATCGCTGGGACACCGGACGAGCTCGGCATCTCGGTGGGCGCTCCCGGACGGAAC 1619
Db 1350 GATCTCTGTTCCACCGCGGAGCGGCGGACACGCGCGCGGTGCTCCGGGACATCTG 1409
QY 1620 CGGCGACACGAATCGGCAAGTACGAGCAGACCTGTACGTGACCGAGGACCTCCGA 1679
Db 1410 GTCCGGCAGCGCGCGCGAGTCTGTGTC-----GTCCGCGGAGCGGCACTCCGAA 1463
QY 1680 CCTCGCGGTACTCGCTGCGCCCTACTTGGGAGCGCGCGCCCTCGACCCCGCGGC 1739
Db 1464 CCCCAGGCGACCGCTCGTGGCGAGCGGCAAGCCCTCCAGCGCAACTTCTTTCTGTG 1523
QY 1740 CCGATCCGTGACCTGCGCATCTCTGTCGACACCGAGCGCTCGAGGTCTTCGTCACGC 1799
Db 1524 GAGGGCGACACCTGCTGTAACCTCTCGAGCGGACCCACCTCGACAGTACGGCACCTC 1583
QY 1800 CGGCGACACCTGCTCTCCAGCAGGTCCACTTCGCGAGGCGGACACGGGAATCTCGCT 1859
Db 1584 GGGCGACA-----CCCGCTGCTCACCGGCTCCGGCTCGGCTCCACACAGGCGAC 1634
QY 1860 CTACACGACGCGCGCGCGGACACTTACCGGCACTCTGTCGCGGAGATTGGCCAGGC 1919
Db 1635 CAAAGCCACCGGCTCTTGGCGGCGAGATCTCTCGGACTTGGCGGAGGCTGCTG 1694
QY 1920 GATCTAGCGGATGCACACCACTGCTACCGAAGCCCGCGCGCGGAGACGCGGCGCA 1979

Db 1695 GCGACGTCGAAACACACGCGCCCTCGCATCTACTCCACCCCTACGACACGACACCG 1754
QY 1980 CAATCGACACGTCCTCTGTCG 1999
Db 1755 CATCAGACCCCTCTCTCCAG 1774
RESULT 7
US-09-476-242-7
; Sequence 7, Application US/09476242
; Patent No. US20020146683A1
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: HARTOG, Karin
; APPLICANT: MARTIN, Eric
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
; FILE REFERENCE: 1605.002
; CURRENT APPLICATION NUMBER: US/09/476.242
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2334
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Leu122-Ser199
US-09-476-242-7

Query Match 4.0%; Score 79.2; DB 10; Length 2334;

Best Local Similarity 43.8%; Pred. No. 2.4e-09;

Matches 458; Conservative 0; Mismatches 573; Indels 15; Gaps 2;

QY 795 COTCTACTGTGTCGACGCGGTTTACCGTTTACCGCCCTGCGCGGACCGGTCATCGT 854
Db 147 CTTGTCTTCCGCGGACGCGCAAGCCTACGACACCGAGGTGCACAACTGTGGCCAC 206
QY 855 CAACACGACGGTCCGCGCGCACACCGCCCGGACATCGAGACGCGGAGTGTTCG 914
Db 207 CCACGCTTGTGCGCCACCGACCCCAACCGCCGAGGATCGTGTGGAGACGTGACCGA 266
QY 915 CGACCCCAAGATCCACTGGGACACCGCCCGGAGAAATGGTCTCGTCTCATCGGACGACT 974
Db 267 GAATTCACATGTGGAAGAACACATGTTGGAGCAGATGACGAGGACATCATCAGCT 326
QY 975 CGGTTACGCGGTTTACACCTTCGCGCAACCTGCGGACTGGACATTCGCGCAACTT 1034
Db 327 GTGGACACAGCGCTGAAGCCTGCTGAAGCTGGGCAACAGCGTGTATCACCCAGGCTG 386
QY 1035 CGACTACCGGAACACCGCCCTCGGGGCGATCGAGTCCCGGACCTGTTTCGAGATCACCG 1094
Db 387 CCCCAGGTGAGTTTCGAGCGCCATCCCATCTACTGCGCGCCCGCGGCTTCGCCA- 445
QY 1095 AGACGACGGGACACCGCCACTGGGTGCTCGCGCCAGCATGGACGCTACGGCATCGGCT 1154
Db 446 -----TCTGTAAGTGCACAGCAAGAAAGTTCAACGCGAGCGCCCTGCAACAGTGA 500
QY 1155 CCCCATGACTAGCGCTTACTGAGAGGACCTTGGACGCGGAGAGTTCACGCGCGACGA 1214
Db 501 CACCGTCAGTGCACCCACCGCATCCGCCCGGTGTGAGCACCACCGCTGCTCTCAACGG 560
QY 1215 CTTACCCCGCAATGGCTTCGACTGGGGCTGGGACTGGTACGGGCGGCTACCTTGCCATC 1274
Db 561 CAGCCTGCGCGAGGAGGCGGTGGTATCCGACGAGAGAACTTACCGCAACAGCGCAAGAC 620
QY 1275 GATCGACGCGCGGACGCAAGCGCTCGCCATCGCGTGGATGAACAACTGGAAGTACGC 1334
Db 621 CATCATCTGTGAGCTGAAGGAGCGGTGGAGTCACTGACCCGCCCAACACAAACAC 680
QY 1335 CGCAGCGACGTCCCGACCGGACGATCCGAGGCTTACACGGGCGGAGAACTCGATCTCGG 1394
Db 681 CCGCAAGAGCATCACCATCGCGCCCGCGGCGCTTCTACGCCACCGGCGACATCATCGG 740

[illegible]

RESULT 8

```

US-09-860-846-32
; Sequence 32, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-32

```

	Query Match	3.9%	Score 78.2	DB 9	Length 11220
	Best Local Similarity	41.9%	Prd. No. 3.4e-09		
	Matches 601	Conservative	0	Mismatches 823	Indels 9
					Gaps 2
Qy	469	GGCTCGCTCCGTCGTACACATGACGCCGCCGAGCGCTGGCTCGGACCCCCAA	528		
Db	2740	GACTCCCCCGGCTCCAGGGCTCGGTGAGACTCCTGGCGCTACCGCATCGACTGGAAG	2799		
Qy	529	CGCCGGTTCACACCCACCGGCCCTACCGTGTACTACCTCGACACGAAACAAC	588		
Db	2800	CGCTCGGGTCGCCGACGCGTCCGAGCGSCGGGCTGTCGGGCGCTGCTGTCGTC	2859		
Qy	589	GGCCCCGGCGCTGGGACCACCGCAGCAGCACCGCGCTCGCTTCACGACACCGGC	648		
Db	2860	GTCCCCGAGGACCGTTCCGCCGAGGCCGCCCGGTGTCGCCCGCGCTGTCCGGCGCCGC	2919		
Qy	649	ACCGTGATGCGCGTCGCGGCCCGACGCTCCCGCGTGTGGTCCGGGTGTCGGCGGCTCGTCGGCACC	708		

Db 4000 GTGACCTGGTGGCCCTGGAGCCCTGGGAGGCGACGCCGGTCAACGAGGGTGGGCGGCGG 4059
QY 1780 GTGAGGCTTTCGTAACCGCGCGCCACACCGGTCTCCAGCAGGTCCACATTCGCGGAG 1839
Db 4060 GAGCGGCTGCGCGCTGGCGCTGGCCCTCGCCCGCGGACGCGCTCAACCGCCCTG 4119
QY 1840 GCGCACGGGAATCTCGCTCTACACCGACGCGCGCGCGGACATTCACCGG 1892
Db 4120 GACACCGGCTCGGCGACGCGGACACCGCGCTCAGCATCGCGGAGTGCAGCTG 4172

RESULT 9

US-09-988-384B-32

; Sequence 32, Application US/09988384B

; Publication No. US20030073824A1

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.536US1

; CURRENT APPLICATION NUMBER: US/09/988.384B

; CURRENT FILING DATE: 2001-11-19

; PRIOR APPLICATION NUMBER: PCI/US99/14398

; PRIOR FILING DATE: 1999-06-25

; PRIOR APPLICATION NUMBER: US 09/105,537

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 53

; SEQ ID NO 32

; LENGTH: 11220

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-988-384B-32

Query Match 3.9%; Score 78.2; DB 9; Length 11220;
Best Local Similarity 41.9%; Pred. No. 3.4e-09;
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;

QY 469 GCGTCCGTCGCTGACATGACGCGCGCGCGCTGCTGCGACGCCCAA 528
Db 2740 GACTCCCGCGCTCAGGGCTCCGTGAGGACTCTTGGCGCTACCGACTGGAG 2799
QY 529 CGCCCGGTACACACCGCGCGCTTACAGGTGTAATCTACTGCACTCCGACCAACAAAC 588
Db 2800 CCGCTCGCGGTGCGGAGCGCTCCGAGCGCGCGGCTGTCGGGCGCTGCTCGTC 2859
QY 589 GCGCCGCGGCTGGGACACGAGACGACGCGGCGTCCGCTTACGACACCGGC 648
Db 2860 GTCCCGGAGGACGTTCCGCGGAGCGCCCGCGGTGCTCGCGCGCTGTCCGCGCGCG 2919
QY 649 ACCGTGATCCGCTCGCGCGGCTTCCCGGTGCTGCGGCTCGCGGCTCGTGGCAC 708
Db 2920 GCGGACCCGTACACTGGAGCTGTCCCGGTGCGGCGACGCGGCTCGCGCGGAG 2979
QY 709 GGAACACGCG---AGGTTTGGCGCGCGCGGCTGCTGCGCTCGACCGACCGCGAC 765
Db 2980 CTGGCGGAGGCGCTGGCGCGCGCGGTGGAGCGCTCGAGCGGCTCTCTCGTCTGTCG 3039
QY 766 GACGCGTCCGCAAGTACAGGACGAGTACCTTACTTACTGTCGACCGCGGCTTACG 825
Db 3040 TGGGACGAGAGCGCGACCCCGCGGACCCCGCGCTTACCCGCGGCGACCGCGCGCAC 3099
QY 826 TTACCGCGCTCGCGGACCGCGCTATCTCAACACCGGTCGCGCGCGCGCACCGGCC 885
Db 3100 CTCACCTGGTGCAGCGCTGGAGACCGCGCGTCCGCGCGCGCTGTGGTGGTGAC 3159
QY 886 GCGGAGATCGAGACCGCGGAGTGGTTTCGCGGACCGCGCGGACATCCACTGGGACACCGCGCG 945
Db 3160 CACGCGCGGCTGCTGCGTGGCGCGCGCGGCGCGGACCGCTCACTCCCGCGCGCGCGCTGTG 3219
QY 946 GGAGAATGGGTCTGCGTCACTCGGAGACTCGGCTACGCGGCTTCTACACCTCCCGGAC 1005

Db 3220 TGGGCGATGGGCGGCTCGCCGCCCTTGAGAGCACCCCGAGCGGTGGGCGGCGCTGATCGAC 3279
QY 1006 CTGCGGAGCTGACACTTTCGCGCAACTTCGACTACCCGAAACACCGCCCTCGCGCGCATC 1065
Db 3280 CTGCGCTCGGAGCGCGGCGCGCTTGAGACCGCATGACCACTGCTCTCCCGCGGCT 3339
QY 1066 GAGTCCCGCGGCTTCTGAGATCACCGAGACGAGGAGACAGGACACTGGGTGCTCGGC 1125
Db 3340 ACGGCTGAGGACGAGTCCGCTAGCGCTCCGGGCTCTCGCGCGCGCTCTCGCTCGCG 3399
QY 1126 GCGACATGGAGCGCTACGGCATCGGCTCCCATGAGTACGCTACTTGGACAGGAC 1185
Db 3400 GCTTCCCTCCCGCGCACGCGACGCTTCGCGCTGTGTCAGCGCGACGCGCATGCTGTC 3459
QY 1186 TGGGACGCGGAGCAGTTCACCGCGGACGCTCACCGCGCAATGGCTCGACTGGGCTGG 1245
Db 3460 GTCACCGGTGCGGAGGAGCTTCGCGCGCGCGGAGCGCGCTGCGCGCGGACGCG 3519
QY 1246 GACTGTAGGCGCGCTCACTTGGCCATGATGAGGCGCGCGGACGCGGACGCGCTCGGC 1305
Db 3520 GCGGACACCTCTCTCTCCACACCGCGCTCCGCGAGCGGCGCGGCGCGGAGCGACCTC 3579
QY 1306 ATCGCGTGGATGAACAACCTGGAAGTACCGCGGACGCTCCCGCACGCGCATCCGAC 1365
Db 3580 GGTGCGCGGAGGACTTCGCGCTCCCGGCTGTCGCGGAACTCGCGGACTGGGCGG 3639
QY 1366 GGCTACAAACGGGCGAGACTCGATCTCGCGGAGCTCGCGCTCGCGCGACGCG-----CT 1419
Db 3640 ACGGCGACCTGCTGACCTCGGACCTCAGGACGCGGAGCGCGCGCGCTGCTCGGC 3699
QY 1420 GCGCGCTGATACACCTCTGAGCACCGCGGTGGAGCGCTGACGAACTACTACCGCG 1479
Db 3700 GCGGTCTCGGCGCGACCGCTCAGCGCGCTCCACCTCGCGCGCGCGCTCGACTCC 3759
QY 1480 ACCACACACTCCCGACGCGGACCTCGACGCGCGCGCTCGCGCTCTGCCATGGAAACGCG 1539
Db 3760 GAGCGCTCGCGCGGACGCGGACGCGCTCGCGCGCTGTCGTGACCGCGGAGCGCAC 3819
QY 1540 GCATACGAGATGAGCTCGATCGCTTGGGACACCGGACGAGCGTGGCATCTCGGTG 1599
Db 3820 GCGCGCTCCACCTGAGCGCTCTCGGAGCGCGGCTGCGGAGCGCTCGCGCG 3879
QY 1600 GCGCGCTCCCGGAGGAAACCGGACAGGAACTCGGAGTACGGAGCAGACCTGTAC 1659
Db 3880 GTCTCTGCTCTCTCTCTCGGTCCCGGATCTGGGCGCGCGGCTCAGGCGCGCTAC 3939
QY 1660 GTGACCGGAGGACCTCGACCTCGCGGCTACTCGCTCGCGCGCTACTCGCGAGCGCG 1719
Db 3940 GCGCGGTACGCGCTTCTCTGACCGCTCGCGGTCAGCAGCGGCGCGCGCGCGCG 3999
QY 1720 GCGCGCATCGACCGCGCGCGGATCCGTGACCTCGCATCTCTCGACACCGAGCG 1779
Db 4000 GTGACCTCGGTGGCGCTGGAGCGCGCTGGGAGGCGACCGCGTACCGGCTCGGACGCG 4059
QY 1780 GTGAGGTCTTGTCAACCGCGCGCGCACACGCTGCTCTCCAGCAGGTCCACTTCCCGGAG 1839
Db 4060 GAGCGCTGCGCGCGCTGCGCTGCGCGCGCTGCGCGCGCGCGGCGCGCTCAGCGCGCTG 4119
QY 1840 GCGGACGCGGAATCTCGCTCTACACGCGCGCGCGCGCGCGCTTACCGG 1892
Db 4120 GACCGCGCTCGGCGCACGCGGACACCGCGCTCAGCATCGCGGAGCTGACTG 4172

RESULT 10

US-09-861-289-32

; Sequence 32, Application US/09861289

; Patent No. US20020110897A1

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin


```
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861.289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Query Match      3.9%; Score 78.2; DB 10; Length 11220;
Best Local Similarity 41.9%; Pred. No. 3.4e-09;
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;

QY 469 GCCTCGCTCCGCTGCTACCATGACGCCCCCGACGGCTGGCTCTGCGACCCCA 528
Db 2740 GACTCCCCCGCGTCCAGGGCTCGGTGACGAGTCTCTGGCGCTACCGCATCGACTGGAAG 2799
QY 529 CGCCCGGTACACACACCGCGCTTACCAGCTGTACTACCTGCACTCCGACACAGAAC 588
Db 2800 CGCTCGGTGCGCGAGCGTTCGAGCGCCGCGGTGTCGCGCGCTGCTGCTGCTC 2859
QY 589 GCGCCCGCGGCTGGGACACGAGCAGCAGCGCGCTGCGCTTACGACACACGCGC 648
Db 2860 GTCCCGAGGACCGTTCGCGCGAGCGCGCGCGTGTCTGCGCGCTGTCTGCGCGCGC 2919
QY 649 ACCGTGATCGCGTGGCGCGGCTTCCCGTGTGTCGCGGTGCGGGTGTGCGGAC 708
Db 2920 GCCACCGCGTACAGCTGTGAGCTGTCCCGCTGGCGGACCGCGAGCGGCTGCGCGGACG 2979
QY 709 CGAACACGCGC---AGGTTTCGCGCGCGCGCTGCTGCGGTGCGGACCGCGGAC 765
Db 2980 CTGGCGAGGCGCTGGCGCGCGCGGTGAGCGGTGCGGCGTCTGCTGCTGCTGCG 3039
QY 766 GACGGCTCCCAAGTACAGGAGCAGTACTTCTACTGGTGCAGCGGCGGGTTCAG 825
Db 3040 TGGACGAGAGCGCACCGCGCGCACCGCGCGCTTACCGCGGCGCACCGCGCGCAC 3099
QY 826 TTCACCGCGCTGCGCGACCGCGTATGTTAAGACCGAGGTGCGCGCGCGCACCGCGC 885
Db 3100 CTCACCTGTGTGAGGCGCTGGAGAGCGCGCGCGTGTGCGCGCGCTGTGCTGCTGAC 3159
QY 886 CGCAGATCGAGAGCGCGAGTGTTCGCGACCGCGCGTTCAGTCCACTGGGACACCGCGC 945
Db 3160 CACGGCGGTGTCGTGCGCGCGCGCGCACCGCTACCTCCCGCGCGCGCGCGCGCG 3219
QY 946 GGAGATGGGTCTCGTATCGGACGACTGCGGTACGCGCGGTTTCTACACCTCCCGAAC 1005
Db 3220 TGGGCGATGGCGCGGTGCGCGCGCTGGAGCACCGCGCGGTGGCGCGCGCTGATCGAC 3279
QY 1006 CTGCGCGACTGAGACTTCGCGCGCACTTCACTACCGGACCGAGCGCTCGCGCGGATC 1065
Db 3280 CTGCGCTCGGACCGCGCGCGCGCGCTTGAACCGCATGACACCGGTCTCTCGCGCGCGGT 3339
QY 1066 GAGTGCCTCCGACCTGTTGAGATCACCGCAGACGCGGACACGCGCTGGGTGCTGCGC 1125
Db 3340 ACGGCTGAGGACAGTGTGCGGTACGCGCTTCGCGGTGCTGCGCGCGCGCTGCTGCGC 3399
QY 1126 GCCAGCATGAGCGCTACGGCATCGGCTCCCGATGAGTACGCTACTGAGAGGACAC 1185
Db 3400 GCCTCCCTCCCGCGCGACGCGCGCTTTCGCGGTGGCGAGCGCGCGCGCGCGGTGCTC 3459
QY 1186 TGGAGCGGAGAGTTCACGCGCGAGACCTACCGCGCAATGGTCTGACTGGGGTGG 1245
Db 3460 GTCACCGGTGCGAGGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3519
QY 1246 GACTGTGAGCGCGCTGCTGCTGCGCATGATCGACGCGCGCGCGAGACCAACGCGCTGCC 1305
Db 3520 GCCGAGACCTCTCTCTCCACACCGCGCTTCCGGCAGGAGCGCGCGCGCGCGCGCTCC 3579
```

RESULT 11

US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US2002016742A1

GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.

APPLICANT: Liu, H.

APPLICANT: Xue, Y.

APPLICANT: Zhao, L.

FILE REFERENCE: 600.438US1
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

CURRENT APPLICATION NUMBER: US/09/860.846

CURRENT FILING DATE: 2001-05-18

PRIOR FILING DATE: 1998-06-26

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 5

LENGTH: 36778

TYPE: DNA

ORGANISM: Streptomyces venezuelae

US-09-860-846-5

Query Match 3.9%; Score 78.2; DB 9; Length 36778;

Best Local Similarity 41.9%; Pred. No. 3e-09;

Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;

QY 469 GCCTCGCTCCGCTGCTACCATGACGCCCCCGACGGCTGGCTCTGCGACCCCA 528

Db 18427 GACTCCCCCGCGTCCAGGGCTCGGTGAGGACTCTCTGGCGCTACCGCATGCTGAAG 18486

QY 529 CGCCCGGTACACACCGCGCGCTACCATGCTGTACTACTGCTGCTGCGGACGAGAAC 588

Db 18487 GCGCTCGGGTCCCGACGCGTCCGAGCGCGCGGGCTGTCGGGGCTGGTCTCGTCTC 18546
QY 589 GCGCCCGGGGCTGGGACACCGAGCAGCAGCGCGCGTGCCTTACGACACACCGC 648
Db 18547 GTCCTCGGAGGACGTTCCGCGAGCGCGCGCGGTGCTCGCGCGCTGTCCGCGCGCGC 18606
QY 649 ACCGTGATGCGCGTCCGGCGCGACATTCCTCCGTTGGTGGTGGGTCGGGGTCTCGGCGAC 708
Db 18607 GCGACCCCGTACAGCTGAGACGTGCTCCCGCTGGCGAGCCGCGAGCGGTCTCGCGCGAG 18666
QY 709 GGAACACCGC---AGGTTGGCGCGCGCGCGGTGCTGGCGTCCGACACGCGGAC 765
Db 18667 CTGGCGAGGCGCTGGCGGCGCGGTGAGCGCTCGAGCGGGTCTCTCGTCTCGCG 18726
QY 766 GACGCGTCCCGAAGTACAGGACGAGTACCTTACTGTGTGAACGAGCGGGTTCACG 825
Db 18727 TGGGAGGAGCGCGACCCCGCGCACCCCGCGCTTACCCGGGGACCGCGCGAC 18786
QY 826 TTCACCGCGCTGCCGACCGCGTCACTGTCACACGAGGTTCGCGCGCGCACCGCC 885
Db 18787 CTCACCTGTGTCAGCGCTGGAGACCGCGCGTCCGCGCGCGCTGTGTGCTGAC 18846
QY 886 GCGGAGATCGAGACCGCGAGTGTTCGCGACCCCAAGATCCACTGGGACACCGCGC 945
Db 18847 CACGCGCGGTGTCGTGCGCGCGCGCGACCATGTCACCTCCCGCGCGCGCATGTG 18906
QY 946 GAGAAATGGGTCTGCTCATCGAGACTGCGGTGACGCGGTTCACACTTCGCGGAAC 1005
Db 18907 TGGGATGCGCGGTGCGCGCTTGAGACACCGCGAGCGGTGGCGCGCTGTATCAG 18966
QY 1006 CTGCGCGACTGAGACTTCGCGCGCACTTCGACTACCGCAACCGCGCTCGGGCGATC 1065
Db 18967 CTGCGCTCGGACCGCGCGCGCGCTTGGACCGCATGACACCGTCTCGCGCGGT 19026
QY 1066 GAGTCCCGCGCTGTCGAGATCACCGACGAGCGGACGCGACCTGGTGTCTCGC 1125
Db 19027 ACGGTGAGGACGAGTTCGGGTACGCGCTCCGGGCTGTCCGCGCGCGCTCTCGC 19086
QY 1126 GCCAGATGAGCGCTACGCGATCGCGTCCCGATGAGTACGCTACGCTACGAGCGAC 1185
Db 19087 GCTCCCTCCCGCGCGCGCGCGCTTCGCGGTGGGAGCGCGCGCGCGCGTGTCTC 19146
QY 1186 TGGGACGCGGAGCTTCACCGCGCGCGCGCTACCGCGCGATGCTGCTGAGTGGGCTG 1245
Db 19147 GTCACCGGTGCGGAGGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19206
QY 1246 GACTGTACGCGCGCTCACCTGGGCTATCGATCGAGCGCGCGCGCGCGCGCGCG 1305
Db 19207 GCGGACACTCTCTCTCCACACCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCG 19266
QY 1306 ATCGGTGGATGAACACTGGAAGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1365
Db 19267 GGTGCGCGGAGGACTCGCGGCTCGCGCGGTTCGTCGCGGAATCGCGGACTGGGCG 19326
QY 1366 GGTACACGCGGAGACTCGATGTCGCGGAGTTCGCGGTGCGCGCGCGCGCGCG 1419
Db 19327 ACGGCGCGGTGTCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19386
QY 1420 GCGCGCTGTACACCTCTCTGAGACCGCGGTGGCGCGGTGACGAATGCTACCGCG 1479
Db 19387 GCGGTCTCGGAGCGCGCGCTAGCGCGCTCTCCACTTCGCGCGCGCGCGCGCG 19446
QY 1480 ACCACCACTTCG 1539
Db 19447 GAGCGCTCG 19506
QY 1540 GCATACGAGATCGAGTGCATCGCTGGGACCGCGCGCGCGCGCGCGCGCGCGCG 1599
Db 19507 GCGCGCTTCCACTGAGCGCGCTCTGCGGAGCGCGCGCGCGCGCGCGCGCGCG 19566
QY 1600 GCGCGCTCCCG 1659

Db 19567 GTCCTGGTCTCTTCTCTCTGTCGCGCGGATCTGGGGCGCGCGGTACAGGCGCGTAC 19626
QY 1660 GTCGACCGAGGACCTTCGACCTCGCGGGTACTCGCTGCGCGCGCTACTCTCGGAGCGCGC 1719
Db 19627 GCGCGCGGTACGCGCTTCTCTGACGCGCTTCGCGGGTCAAGACCGCGCGCGCGCGCG 19686
QY 1720 GCGCGCGGTACG 1779
Db 19687 GTGACCTCGGTGCGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19746
QY 1780 GTCGAGGTCTTCTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1839
Db 19747 GAGCGGTGCGCGCGCGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19806
QY 1840 GCGGACACGGAATCTCGCTTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1892
Db 19807 GACACCGCGCTCG 19859

RESULT 12

US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match 3.9%; Score 78.2; DB 10; Length 36778;
Best Local Similarity 41.9%; Pred. No. 3e-09;
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;

QY 469 GCGTCCGCTCCGTCGCTTACACATGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 528
Db 18427 GACTCCCGCGCGTCCAGGGTCCGTGCGAGGACTCTTGGCGCTACCGCATCGACTGGA 18486
QY 529 GCGCGGTACACACCGCGCGCTTACAGTGTACTACCTGACCTCCGACCGACAGAAC 588
Db 18487 GCGCTCGCGGTGCGCGCGCGTCCGAGCGCGCGCGCGCGCGCGCGCGCGCG 18546
QY 589 GCGCGCGCGGTGCGGACACGCGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 648
Db 18547 GTCCCGCGAGGACCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18606
QY 649 ACCGTGATGCGCGTCCGCGCGCGCGCTTCCCGGTGCTGGTGGTGGCGGTCTGTCGGCAC 708
Db 18607 GCGGACCGGTACAGTGGAGGTGTCGCCGTGGGCGAGCGCGCGCGCGCGCGCG 18666
QY 709 GCGAACACGCG---AGGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 765
Db 18667 CTGGCGAGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18726
QY 766 GACGCGTCCCGAAGTACAGGACGAGTACCTTACTGCTGACCGCGCGCGGTTCACG 825
Db 18727 TGGGACGAGAGCG 18786
QY 826 TTCACCGCGCTGCCG 885
Db 18787 CTCACCTGCTGACGCGCTGGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18846

Y 886 GCCGAGATCGAAGCCGCGAGTGGTTCCGCGACCCCAAGATCCACTGGGACACCCCGC 945
D 18847 CACGCGCGGTGTCTCGTCGGCGCGGACACAGGTCACTCCCGCGCCAGGCCATGGTG 18906
Y 946 GGAGATGGGTCTGGTTCATCGGAGGACTGGGTACGGCCGGTTCATACACCTCGCCGAC 1005
D 18907 TGGGCGATGGCGCGGTCTCGCCCTCGAGACCCCGAGCGGTGGGGCGGTGATCGAC 18966
Y 1006 CTGCGCGACTGGACACTTCGCGCGCAACTTCGACTACCCGGAACACAGCCCTCGGGGCGATC 1065
D 18967 CTGCGCTGGAGCGGACCGCGCGGCTGGACCGCATGACACAGTCTCGCGCGCGGT 19026
Y 1066 GAGTCCCCCGACCTCTTTCAGATCACCGAGACGAGCGGACGCACTGGGTGCTCGCC 1125
D 19027 ACGGTGAGGACAGGTGCGCGGTACGCGCTTCGCGCGGTGCTCGCGCGCGCTCGTCCG 19086
Y 1126 GCCAGCATGGAGCGCTACGGCATCGGCTCCCGCATGAGCTAGCGCTACTGACAGGAC 1185
D 19087 GCTTCCCTCCCGCGGCGACGCGACCGCTTCGCGGTGGTGGCAGGCGGACGCGACGTGTC 19146
Y 1186 TGGAGCGCGAGCACTTCACCGCGGACGACCTCACCCCGCAATGCTCGACTGGGGCTGG 1245
D 19147 GTCACCGGTGCGGAGGACCTTGGCGCGCGCGGCGGCGGCTGGCGCGGCGGACGCG 19206
Y 1246 GACTGGTACGGCGCGTCACTTGGCCATCGATCGACGCGCGCGGACGCAAGCGCTCGCC 1305
D 19207 GCGGACACCTCTCTCCACACACCCCTCCGCGACGAGGCGCGGAGGCGGCGGACCTCC 19266
Y 1306 ATCGGTGGATGAACAATGGAAGTACGCGCGACGCGGAGTCCCGACGAGCATCCGAC 1365
D 19267 GGTGCGCGCGGAGACTCGCGGCTCGCGGGCTGCTGCGCGCAACTCGCGGAGCTGGGCG 19326
Y 1366 GCTTACAAACGCGGCAACTCGATCGTCGCGAGCTGCGGCTCGCGCGGACGCTCT 1419
D 19327 ACGGCCACCTGCTGACCTCGACCTACGAGCGGAGGCGGCGCGGCTGCTCGCC 19386
Y 1420 GCGGCTGGTACACCTCTTACGACACCCCGTGGAGCGGTGAGCACTACGACCTACCGCC 1479
D 19387 GCGGTCTCGGAGCGACCGCTAGCGCGCTGCTCCACCTGCGCGCGGCGGCGGCTGCT 19446
Y 1480 ACCACCACTTCCCGACCGGACCGTTCGACGCGGCGGCGGCTGCTGCGATGAGGAGCG 1539
D 19447 GAGCGCTCGCGCGGACGCGGAGCGGCTGCGCGGCTGCTGCGGCGGCGGCGGCGG 19506
Y 1540 GCATACGAGATCGAGCTCGACATCGCTGGAGACCGCGGACGAGCTCGGATCTCGGTG 1599
D 19507 GCGGCTCGACCTGAGACCGCTCTGCGGAGGCGGCGGCTGCGGAGGCGGCTCGCGCC 19566
Y 1600 GCGCGCTCCCGGAGCGGACCGGACACGATCGGCAAGTACGAGGAGGAGCTGATC 1659
D 19567 GTCCTGGTCTTCTCTCTCGTGGCGGATCTGGGCGGCGGCGGCTGAGGCGGCTGAC 19626
Y 1660 GTCGACGAGGAGCTCGGACCTCGCGGGTACTGCTGCGCGGCTGCTGCGGAGCGGCG 1719
D 19627 GCGCGGTAGCGGCTTCTCGAGCGGCTGCGCGGTGCTGAGCGGCGGCGGCGGCGGCG 19686
Y 1720 GCGCGCATCGAGCGGCGGCGGATCGGTCGACCTGCGGATCTGCTGACACCGGAGC 1779
D 19687 GTGACCTGGTGGCTGAGGCGGCTGGGAGGCGGAGCGGCTGACCGAGGCTGCGACCGG 19746
Y 1780 GTCGAGGTCTTCTGCAAGCGGCGGACACCGTGTCTTCCGAGGAGTCCACTTGGCGAG 1839
D 19747 GAGCGGTGCGCGGCTTGGCGGCTGCGCGGCTGCGCGGCGGAGGCGGCTGCGCGGCTG 19806
Y 1840 GCGGACGAGGAGTCTGCTCTACACGAGCGGCGGCGGCGGCGGCGGCTTACCGG 1892
D 19807 GACACCGGCTCGGCGGAGGAGACCGCGCTGACGATGCGCGGAGCTG 19859

RESULT 13

US-09-988-384B-5

; Sequence 5, Application US/09988384B

; Publication No. US20030073824A1

; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536051
; CURRENT APPLICATION NUMBER: US/09/988,384B
; PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5
; LENGTH: 37948
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-988-384B-5

Query Match 3.98; Score 78.2; DB 9; Length 37948;
Best Local Similarity 41.98; Pred. No. 3e-09;
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;
Y 469 GGCTCGCTCCGTGCGCTTACCATATGACGCGCCCGCCAGCGGTGGCTCTGGGACCCCA 528
D 19597 GACTCCCCCGCGTCCAGGGCTCGGTGCGAGGACTCTCGGCGCTACCGCATCGACTGGA 19656
Y 529 GCGCGGTACACACCGGCGGCTTACAGCTGTACTACTGCTGCTGCGACGACGACGAC 588
D 19657 GCGCTCGGCGTGGCGGCGGCTCGGAGCGGCTCGGCGGCTGGCTGCTGCTGCTG 19716
Y 589 GCGCGCGCGCTGGGACACGCGAGACGACGAGCGGCTGCTTACGACGCGGCTTACGAC 648
D 19717 GTCCCGGAGGACGCTTCCGCGAGGCGCGCGGCTGCTGCGGCGGCTGCTGCGGCG 19776
Y 649 ACCGTGATCGCGCTCGGCGCGGCTTCCCGTGTGGTGGTGGTGGTGGTGGTGGTGGTGG 708
D 19777 GCGGACCGGCTACAGCTGGAGCTGTCCCGCTGGGCGACCGGCGGCTGCGGCGGCG 19836
Y 709 GCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 765
D 19837 CTGGCGGAGGCGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 19896
Y 766 GAGCGCGTCCGCAAGTACAGGAGGAGTACCTTACTGCTGACGCGGCGGCGGCTTACG 825
D 19897 TGGGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 19956
Y 826 TTCACGCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 885
D 19957 CTCACCGTGGTGGAGGCGCTGGAGGACGCGGCGGCTGCGGCGGCGGCGGCGGCGG 20016
Y 886 GCGGAGATCGAGACGCGGAGTGGTTCGCGGACCGCGGCGGCGGCGGCGGCGGCGGCG 945
D 20017 CACGCGCGGCTGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 20076
Y 946 GGAGATGGGTCTGGTTCATCGGAGCTGCGGTACGCGGCGGCGGCGGCGGCGGCGGCG 1005
D 20077 TGGGCGATGGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 20136
Y 1006 CTGCGCGACTGGACACTTCGCGCGGCAACTTCGACTACCGGCGGCGGCGGCGGCGGCG 1065
D 20137 CTGCGCTCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20196
Y 1066 GAGTCCCCCGGAGTGTTCGAGATCACCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1125
D 20197 ACGGTGAGGACGAGTTCGCGGCTACGCGCTCGGCGGCTGCTGCGGCGGCGGCGGCG 20256
Y 1126 GCGGACGAGGCGCTACGCGCTCGGCGCTCCCGCTACGCTGAGCTGAGCTGAGGAGGCG 1185
D 20257 GCTTCCCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20316

QY 1186 TGGGACGGGACAGATTCCACCGCAGACACTCACCCCGCAATGCTCGACTGGGGCTGG 1245
Db 20317 GTACCGGTGCGGAGAGCTCGCGCGCGAGGCGCGACGCGGCTGGCCCGCAGCGC 20376
QY 1246 GACTGGTACGGCGGTCTACCTTGGCCATCGATCGAGGCGCCCGAGACCAAGCCCTGGC 1305
Db 20377 GCGGACACCTCTCTCTCCACACCAACCCCTCCCGCAGCAAGGCGCGCAAGGCACCTCC 20436
QY 1306 ATCCGCTGGATGAACAACTTGAAGTACGCGCAGCGAGCTCCCGACGACGACATCCGAC 1365
Db 20437 GGTGCGCGGAGACTTCGGCCCTCGCGGCTGCTGCGGAACTCGCGGACCTGGGCGGG 20496
QY 1366 GGTCAACAGGCGAGACTCGATCTCGCGAGCTGCGGCTCGCCGACAGC-----CT 1419
Db 20497 ACGGCAACCGTCTGACCTTGCACCTACGGACCTACGGACGCGGAGGCGCGCGGCTGCTCGC 20556
QY 1420 GCGGCTGGTACACCTCTGAGACACCCCGCTGGGAGCGCTGACAACTACCTACCGGC 1479
Db 20557 GCGGCTCGGAGCGGACCGGCTCAGCGCGCTCTCCACCTGCGCGCCACCGTCACTCC 20616
QY 1480 ACCACACACTCCCGGACCGGACCGCTCGACGCGAGCGGCTCTGCCATGGAACGAGCGC 1539
Db 20617 GAGCGCTCGCGGACCGGACGCGCTCGCGCTGCTGCTGACCGCGAAGCCACC 20676
QY 1540 GCATACGAGATCGAGCTCGACATCGCTGGGACACCGGAGCAAGCTCGGCACTCTCGGTG 1599
Db 20677 GCGGCTCTCACTGAGCGGCTCTGCGGAGGCGCGGCTGCGGAGGCGGCTCGCGGCTCGCGCC 20736
QY 1600 GCGCGCTCCCGGAGGACCGGACGACAGACATCGCAAGTACGAGGACGACCTGTAC 1659
Db 20737 GTCTGTGCTCTCTCTCGGTCGCGCGATCTGGGCGGCGGCTGAGGCGCGGTAC 20796
QY 1660 GTGACGAGGACCTCGGACCTCGCGGCTACTCGCTCGCGCCCTACTCGGAGCGCGCC 1719
Db 20797 GCGCGGTGAGCGCTTCTCGACGCTCTCGCGGTGACGACCGGCGGCGGCGCGCCACC 20856
QY 1720 GCGCGGCTGAGCGCGCGCGGCTCGTGCACCTGCGGCTGCTCGACACCGCAGAGC 1779
Db 20857 GTGACCTCGGTGCGGCGCGCTGCGGAGGCGCGGCTGCGGAGGCGGCTGCGGAGGCTGCGAGCGG 20916
QY 1780 GTGAGGCTTGTCAACGCGGCGGACGCGCTCTCCAGAGGCTCCACTTCGCGCGAG 1839
Db 20917 GAGCGGCTGCGCGCTGCGGCTGCGCGCTGCGCGCGGCGGCGGCGGCTGCGCGCGCTG 20976
QY 1840 GCGGACGCGGAACTCTGCTTACACGAGCGGCGGCGCGGCGGCGGCGGCGGCGG 1892
Db 20977 GACACCGCGCTCGGCGCACCGGACACCGCGCTCAGGATCGCGGAGTCTGACTG 21029

RESULT 14

US-09-988-384B-3/c

; Sequence 3, Application US/09988384B

; Publication No. US20030073824A1

; GENERAL INFORMATION:

; APPLICANT: Sherman, D. H.

; APPLICANT: Liu, H.

; APPLICANT: xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600,536US1

; CURRENT APPLICATION NUMBER: US/09/988,384B

; CURRENT FILING DATE: 2001-11-19

; PRIOR APPLICATION NUMBER: PCT/US99/14398

; PRIOR FILING DATE: 1999-06-25

; PRIOR APPLICATION NUMBER: US 09/105,537

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 53

; SEQ ID NO 3

; LENGTH: 12441

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-988-384B-3

Query Match 3.9%; Score 77.6; DB 9; Length 12441;
Best Local Similarity 42.9%; Pred. No. 4.6e-09;
Matches 680; Conservative 0; Mismatches 879; Indels 25; Gaps 5;

QY 367 CCGGCCATCTACGCGCGGCGCTCTCCAGGAGCGGCGGCGGAGGACTCGCCCTGATC 426
Db 11836 CCGGACACCGCAACTCTGACCGGACGAGGTGGCGCGCGGTTCACACCGCGACCTCG 11777
QY 427 TTCGCGGTGTGTGTCGCGCTTCACGCGGCGATCCGCTCGCGGCTCGCTCGCTCGCTC 486
Db 11776 GCGGCTGCGGCTCACACTTCGCGGCGCGCTTCGCGGCGCGCTTCGAGTTCCACGCCA 11717
QY 487 TACCACATGACGCGCGCGCGCTGCTCTGCGACCGCGCGCGCGCGCGCGCGCGCGCG 546
Db 11716 GCGGACGAGCAGCGCTGCGGCTGCTTCGACGCGCGCGCGCGCGCGCGCGCGCGCT 11657
QY 547 GCGGCTTACGAGCTTACTACCTGCTGCGACGAGACGCGCGCGCGCGCGCGCGCGCG 606
Db 11656 GAGCGCGCGCGCGCGCGCGCTGCGGAGCGCTGCGGCGCGCGCGCGCGCGCGCGCG 11597
QY 607 CACGCGAGCAGCAGCGCGCGCTGCTTACGCGACGCGCGCGCGCGCGCGCGCGCGCG 666
Db 11596 GCGGTCACGCGCTTTCGAGGCGCGCGCTGCTTACGCGAGCGCGCGCGCGCGCGCG 11537
QY 667 CCGGACTTCCCGGTGTGTCGCGGCTGCGGCGCTGCTGCGGACGCGCGCGCGCGCGCG 726
Db 11536 ATCCGCGCGCTTCCACAACTTCGCGCTTTCGCGGCGCGCGCGCGCGCGCGCGCG 11477
QY 727 GCGCGCGCGCGCGCTGCTGCGGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 786
Db 11476 AACGCGAAGATGAGCG 11429
QY 787 GAGCAGTACCTTCTACTGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 846
Db 11428 GAGCGGTTTCCGAGGTCTGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11369
QY 847 GTATCTGTAACACGAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 906
Db 11368 CTCGCGGACCTTCCCGCGGCTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11309
QY 907 TGGTTCG 966
Db 11308 CAGTACGTGATGCTGCGAGTTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11249
QY 967 GGACGACTCGGTACGCGCGCTTCTACACCTGCGCGCGCGCGCGCGCGCGCGCGCG 1020
Db 11248 GAGGCTCTGAGGCGCGAGGCGGTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 11189
QY 1021 CTTGCGCGCACTTCTGCTACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
Db 11188 CTGGAGCGGTATCCG 11129
QY 1081 TTCGAGATCAGCGCAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
Db 11128 CCGGTGCTTCTGCG 11069
QY 1141 TACGCGATCGCGCTTCCCGATGACGCTTACCTGCGAGCGCGCGCGCGCGCGCGCG 1200
Db 11068 GACCTGCTGCGTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11009
QY 1201 TTCACGCGCGGAGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
Db 11008 GCG 10950
QY 1261 GTCACCTGGCGATGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320
Db 10949 ACCGCGCGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10890
QY 1321 AACTGGAAGTACG 1380
Db 10889 CTGGGCGCGCGGTCTCAGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10830
QY 1381 AACTCGATCGTCCG---CGAGTGGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCG 1437

Db 10829 GACGAGCCGGAAGACGCGCCGACCTCGTCCGGTTCGGCTACGGCAAGC 10770
Qy 1438 CTGAGCACCCTGCGAGCGCTGACGAATGCTACCGCCACCAACACATCCCGGAC 1497
Db 10769 CCCTTACCCGCTGAGAGGCGCCGACGACCTGGCGGTGACCGGGAGCCCTCCGG 10710
Qy 1498 CGGACCTGCGAGCGAGCGCGCTCTCCCATGGAACGAGCGCATACGAGATCGAGCTC 1557
Db 10709 CGCCTCTCGCCCTGTTCCGGCAGGTCCCGAGCTCCGACCGCGGTGAGACCGGCC 10650
Qy 1558 GACATCGCTGGGACACCGGACGAACTCGGATCTCGGTGGCCGCTCCCGGACGGA 1617
Db 10649 GCGGGGCGTACTGGAAGAACACCTCTCCGCTCGAAGACGCGCGCTTTCGACGG 10590
Qy 1618 ACCCGGCACACGAA--CATCGGCAAGTACGAGACGACACTGTACGTGCGACCGAGACCC 1674
Db 10589 GCGTCCCGAGGAGCGCTTCCGATACGCTGCGGCTCTACCCGCGGCGACCTGC 10530
Qy 1675 TCGACCTCGCGGGTACTCGCTGCGCCCTACTTCGCGAGCGCGCCGCGCCCATCGACCCC 1734
Db 10529 ATGTTCCGCTGCCACTTCTGCTCGGTGACCGCGCGCGCTACGACCGCTCGGCCCTC 10470
Qy 1735 GCGCCCGATCGCTGCGACTCGGATCCTCGTGCACACCGACGCTGCGAGTCTTCTGTC 1794
Db 10469 GACCGCGCAACGCACTTTCGCGTCTGATCGACGAGATACCGCGGGCAACCCCTCG 10410
Qy 1795 AACCGCGGCACACGCTCTCCACAGTCCACTTCGCGAGCGCGCGCGCGCGGAGTAC 1854
Db 10409 GCGATGACTTCTCGCGCGCGCTGGAGCGCTACCAACCGCGCGCTCGGGAGCTGGCC 10350
Qy 1855 TCGCTCTACACCG 1914
Db 10349 GCGACCGCACCGACCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10290
Qy 1915 CAGCGATCTAGCGGATGCACAC 1938
Db 10289 GAGCGCACCTGGAGCG

RESULT 15

US-09-860-846-3/c
; Sequence 3, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U51
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-3

Query Match 3.9%; Score 77.6; DB 9; Length 13613;
Best Local Similarity 42.9%; Pred. No. 4.5e-09;
Matches 680; Conservative 0; Mismatches 879; Indels 25; Gaps 5;
Qy 367 CCGGCCATCTCAGCGCGCGCTGCTCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 426
Db 12523 CCGGACACCGGCAACCTTCGACCGCGGACGAGTGGCGCGCGGTCTACACCCCGACCTCG 12464
Qy 427 TTCGGCGGTCTGTGCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 486

Db 12463 GCGTCTGCTGCGCTGCCACCTCTGGGGCGCGCCCTGCGCGCGCGCGCGCGCGCGCG 12404
Qy 487 TACCACATGACGCGCGCGCGCGCTGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 546
Db 12403 GCGGACGAGCAGCGCTGCGGCTGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 12344
Qy 547 GGGCGCTTACCAGTGTACTACCTGCTGCTGCGACGAGCGCGCGCGCGCGCGCGCGCG 606
Db 12343 GACGGCGCGCGCGCGCGCGCGCTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 12284
Qy 607 CAGCGAGCAGGACGCGCGCGCGCTTACGCGACGCGCGCGCGCGCGCGCGCGCGCGCG 666
Db 12283 GCGCTAACGCTTTCGAGGCGCGCGCTGCTACACGAGCGCGCGCGCGCGCGCGCGCG 12224
Qy 667 CCGGACTTCCCGGTGCTGCGGCTGCGGCTGCTGCGGCGCGCGCGCGCGCGCGCGCGCG 726
Db 12223 ATCGCGCGCTTCCACAACTTTCGCTTTCGACTGCGCGCGCGCGCGCGCGCGCGCG 12164
Qy 727 GGGCGCGCGCGGCTGCTGCGGCTGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 786
Db 12163 AACGCCAAGATGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12116
Qy 787 GAGCAGTACCTTACTGCTGCTGCGAGCGCGCGGTTTACGTTTACCGCGCGCGCGCGCG 846
Db 12115 GACGCGTTTCCGAGGTCATCGACCGGAACCGCGCGCGCGCGCGCGCGCGCGCGCG 12056
Qy 847 GTCATGCTAACACGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 906
Db 12055 CTCGCGGACCTTCCCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11996
Qy 907 TGGTTCCCG 966
Db 11995 CAGTACGTGATGCTGAGATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11936
Qy 967 GGAGCATCGGCTGCGCGCGGTTTACACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
Db 11935 GAGTCTGGAAGCGGAGCGGTGCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11876
Qy 1021 CTTGCGCGCAACTTTCGACTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
Db 11875 CTGAGCGGTACCG 11816
Qy 1081 TTCCGATACCG 1140
Db 11815 CCGGTGCTGCTGCTGCG 11756
Qy 1141 TAGGGATCGGCTTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Db 11755 GACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11696
Qy 1201 TTCCACG 1260
Db 11695 GCG 11637
Qy 1261 GTCACCTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Db 11636 ACCG 11577
Qy 1321 AACTGGAATGAGCG 1380
Db 11576 CTGGGCG 11517
Qy 1381 AACTGATGCTGCG--CGAGCTGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1437
Db 11516 GACGAGCGCGGAGCG 11457
Qy 1438 CTGAGCACCCTGCG 1497
Db 11456 CCCTTACCGCGCTGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11397
Qy 1498 CCGACCTGCGAGCG 1557
Db 11396 CGCTCTCGCGCTGTTTCG 11337

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 02:37:06 ; Search time 5273.5 Seconds
(without alignments)
11042.896 Million cell updates/sec

Title: us-09-868-328b-3
Perfect score: 2001
Sequence: 1 gcgtgcaccccgactccc.....atcgacacgtctctcgctt 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1566	78.3	1566	1	AF181254	AF181254 Arthrobac
2	977	48.8	1890	1	AF294730	AF294730 Microbac
3	945.2	47.2	1752	1	AB001984	AB001984 Arthrobac
4	931.8	46.6	1551	6	E22826	E22826 Levan fruct
5	909.6	45.5	1452	6	E22825	E22825 Levan fruct
6	370.4	18.5	1863	1	AB062062	AB062062 Microbac
7	147.6	7.4	65140	6	AX211705	AX211705 Sequence
8	147.6	7.4	123580	1	AF263912	AF263912 Streptomy
9	147.6	7.4	125401	6	AX211739	AX211739 Sequence
10	141.8	7.1	3030	8	VCA429230	AJ429230 Volvox ca
11	119.8	6.0	3849	6	AX058889	AX058889 Sequence
12	119.8	6.0	12744	1	AF204401	AF204401 Streptomy
13	114.2	5.7	22449	1	SPSNBCDE	Y11548 S.pristinae
14	114.2	5.7	22449	1	SPSNBCGEN	X98690 S.pristinae
15	112	5.6	77457	1	AF210249	AF210249 Streptomy
16	111.8	5.6	2951	1	AF123319	AF123319 Streptomy
17	111.8	5.6	43147	1	SC4A10	AF109663 Streptomy
18	109.2	5.5	27541	6	AX211706	AX211706 Sequence
19	106.8	5.3	30000	6	AX250262	AX250262 Sequence
20	106.8	5.3	88063	1	BSZ94043	Z94043 B.subtilis
21	106.8	5.3	209510	1	BSUB0018	Z99121 Bacillus su
22	106.4	5.3	113193	1	AF357202	AF357202 Streptomy
23	105.2	5.3	15424	1	AE007019	AE007019 Mycobacte
24	105.2	5.3	58930	2	AC098321	AC098321 Rattus no
25	103.8	5.2	47852	1	MTV023	AL022022 Mycobacte
26	103.6	5.2	14860	1	AE007093	AE007093 Mycobacte
27	103.6	5.2	63033	1	MTV008	AL021246 Mycobacte
28	102.4	5.1	303091	2	AC084799	AC084799 Mus muscu
29	102.2	5.1	298166	2	AC087563	AC087563 Homo sapi
30	101.6	5.1	5994	1	SVSNBDE	Y11547 S.virginiae
31	101.4	5.1	42655	1	SC7H2	AL109732 Streptomy
32	101	5.0	15311	1	AE007164	AE007164 Mycobacte
33	99.8	5.0	4068	1	ANU12274	U12274 Actinomycet
34	99.8	5.0	37245	1	SC5F2A	AL049587 Streptomy
35	99.8	5.0	40790	1	MTCY493	Z95844 Mycobacteri
36	99.4	5.0	220469	2	AC074307	AC074307 Mus muscu
37	99	4.9	15348	1	AE007163	AE007163 Mycobacte
38	98.6	4.9	67200	1	MTV017	AL021897 Mycobacte
39	98.6	4.9	77534	1	AF235504	AF235504 Streptomy
40	98.2	4.9	110000	2	LMFLCHR32_06	Continuation (7 of
41	97.8	4.9	34108	1	SCE25	AL354048 Streptomy
42	97.6	4.9	150754	9	AC023491	AC023491 Homo sapi
43	97.2	4.9	10420	1	AE005127	AE005127 Halobacte
44	97	4.8	4615	1	STMWHIB12X	L22864 Streptomyce
45	96.6	4.8	46166	1	SC8F11	AL353864 Streptomy

ALIGNMENTS

RESULT 1
AF181254
LOCUS AF181254 1566 bp DNA linear BCT 02-JUN-2000
DEFINITION Arthrobacter ureafaciens levan fructotransferase (lfta) gene,
complete cds.
ACCESSION AF181254
VERSION AF181254.1 GI:8163731
KEYWORDS Arthrobacter ureafaciens.
SOURCE Arthrobacter ureafaciens
ORGANISM Arthrobacter ureafaciens
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
REFERENCE 1 (bases 1 to 1566)
AUTHORS Song,K.B., Rhee,S.K. and Yoo,E.J.
TITLE Nucleotide sequence of levan fructotransferase gene (lfta) from

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Arthrobacter ureafaciens K-2032
Unpublished
2 (bases 1 to 1566)
Song, K.B., Rhee, S.K. and Yoo, E.J.
Direct Submission
Submitted (27-AUG-1999) Microbial Metabolic Engineering RU, Korea
Research Institute of Bioscience and Biotechnology, Taejeon 305-600,
South Korea

FEATURES
source

1..1566
Location/Qualifiers
/organism="Arthrobacter ureafaciens"
/strain="K-2032"
/db_xref="taxon:37931"
/country="South Korea"
1..1566
/gene="lftA"
1..1566
/gene="lftA"
/function="diffructose dianhydride IV (DFA IV) formation
from levan"
/note="levan-assimilating enzyme"
/codon_start=1
/transl_table=11
/product="levan fructotransferase"
/protein_id="AAF73829.1"
/db_xref="GI:8163732"
/translation="MPLAISRAVLQAGAGALALIFGAVPPAARASAPGSLRAYVH
PTPSSGLCDQPVYTHGAYQLYLHSDNPGGMDHASTIDGVAFTHGTVNPLR
PDPVMSGSVGTANTAGFAGAVVALATQPDGVRYQEQYLWSDYDFTFALP
DPVINTDFQPAHLLGIECPDLFEITADDGTHVWLAASMDAYGLIPMTVAYWTGW
WTLRNFDPNHALGIECPDLFEITADDGTHVWLAASMDAYGLIPMTVAYWTGW
DGEQFADHLLPQWLDWMDYAAVWESIDAPEKRLAIAMNNWYKAAKDVPTDAS
DGYNGNSIVRELRLARQPGWYLLSPVAALTNVATTTLPDRTVDGSAVLDPWG
RAYETELDIATNTAVNGISVSRSPDTRHTNIGRYGADLYVDGSPSLAGYS
RAAAPIDPGARSVHLRLVDTQSFVEFVNAGHTVLSQOVHFAEGDTGSLVTDGPAH
FTGIIVVREIGQAI"

BASE COUNT 268 a 635 c 458 g 205 t

ORIGIN

Query Match 78.3%; Score 1566; DB 1; Length 1566;
Best Local Similarity 100.0%; Pred. No. 1.1e-160;
Matches 1566; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 ATGACCGCGGCATCTACGCGCGCGCGTCTCCAGGAGCGCGCGCGGAGACTCGCC 420
DB 1 ATGACCGCGGCATCTACGCGCGCGCGTCTCCAGGAGCGCGCGCGGAGACTCGCC 60
QY 421 CTGATCTCGGGGTCTGTGGCGCTGAGCGCGCGGATCGCTCCGGGCTCGCTCCGT 480
DB 61 CTGATCTCGGGGTCTGTGGCGCTGAGCGCGCGGATCGCTCCGGGCTCGCTCCGT 120
QY 481 GCGGTCTACACATGACGCGCGCGCGTCTGCGACCGCGCGCGCGCGCGCGCGCG 540
DB 121 GCGGTCTACACATGACGCGCGCGCGTCTGCGACCGCGCGCGCGCGCGCGCGCG 180
QY 541 ACCGCGCGCGCTACGAGTGTACTACCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
DB 181 ACCGCGCGCGCTACGAGTGTACTACCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 601 TGGGACACGCGAGCAGCAGCGCGCGCTTCCACGCGCGCGCGCGCGCGCGCGCG 660
DB 241 TGGGACACGCGAGCAGCAGCGCGCGCTTCCACGCGCGCGCGCGCGCGCGCGCG 300
QY 661 CTGCGCGCGCGCTTCCCGGTGTGGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
DB 301 CTGCGCGCGCGCTTCCCGGTGTGGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
QY 721 GGGTTCGCGCGCGCGCGGTGTGGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
DB 361 GGGTTCGCGCGCGCGCGGTGTGGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
QY 781 TACAGGAGCAGTCTCTACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840

DB 421 TACAGGAGCAGTACTCTCTGTTGTCAGGCGGGGTTTACGTTTACCCTGCCCC 480
QY 841 GACCCCGTTCATCTCAACACCGAGCGGTGCGCGCGCGCGCGCGCGCGAGATCGAGAAC 900
DB 481 GACCCCGTTCATCTCAACACCGAGCGGTGCGCGCGCGCGCGCGCGCGAGATCGAGAAC 540
QY 901 GCGGAGTGTTCGCGAGCCCAAGATCCACTGGGACACCGCGCGCGCGAGAAATGGGTCTCG 960
DB 541 GCGGAGTGTTCGCGAGCCCAAGATCCACTGGGACACCGCGCGCGCGAGAAATGGGTCTCG 600
QY 961 GTCATCGGAGCAGTCTCGGTACGCGTTCACCTCGCGCGCGCGCGCGCGCGCGCGCG 1020
DB 601 GTCATCGGAGCAGTCTCGGTACGCGTTCACCTCGCGCGCGCGCGCGCGCGCGCGCG 560
QY 1021 CTTGCGCGCAACTTCGACTTACCCGAACACCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
DB 661 CTTGCGCGCAACTTCGACTTACCCGAACACCGCGCGCGCGCGCGCGCGCGCGCGCG 720
QY 1081 TTTGAGATCACGCGAGACCGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
DB 721 TTTGAGATCACGCGAGACCGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
QY 1141 TACGGCATCGCGCTCCCATGACGTACGCGTACGTGACAGGCGCGCGCGCGCGCGCGCG 1200
DB 781 TACGGCATCGCGCTCCCATGACGTACGCGTACGTGACAGGCGCGCGCGCGCGCGCGCG 840
QY 1201 TTTACGCGCGAGCAGTCTACCCCGCAATGGTTCGACTGGGGGTGGGACTTGGTACGCGGCG 1260
DB 841 TTTACGCGCGAGCAGTCTACCCCGCAATGGTTCGACTGGGGGTGGGACTTGGTACGCGGCG 900
QY 1261 GTCACTGGCGCATCGATCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320
DB 901 GTCACTGGCGCATCGATCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
QY 1321 AACTGGAAGTACG 1380
DB 961 AACTGGAAGTACG 1020
QY 1381 AACTCGATCGTCCGCGAGTCTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1440
DB 1021 AACTCGATCGTCCGCGAGTCTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
QY 1441 AGCACCCCGTGGCAGCGCTGACGAACTACGTCACGCGCGCGCGCGCGCGCGCGCGCGCG 1500
DB 1081 AGCACCCCGTGGCAGCGCTGACGAACTACGTCACGCGCGCGCGCGCGCGCGCGCGCG 1140
QY 1501 ACCGTGCGAGCGCGCGCGCTGCGCATGGAACGAGCGCGCGCGCGCGCGCGCGCGCGCG 1560
DB 1141 ACCGTGCGAGCGCGCGCGCTGCGCATGGAACGAGCGCGCGCGCGCGCGCGCGCGCGCG 1200
QY 1561 ATCGCTGGGACACCGCGAGCAACGTCGCGCATCTCGGTGGCGCGCTTCCCGCGCGCGCG 1620
DB 1201 ATCGCTGGGACACCGCGAGCAACGTCGCGCATCTCGGTGGCGCGCTTCCCGCGCGCGCG 1260
QY 1621 CGGACACGAAATCGGCAAGTACGAGGAGAGCGTGTAGTCGCGCGCGCGCGCGCGCGCGCG 1680
DB 1261 CGGACACGAAATCGGCAAGTACGAGGAGAGCGTGTAGTCGCGCGCGCGCGCGCGCGCGCG 1320
QY 1681 CTCGCGGGTACTCGCTCG 1740
DB 1321 CTCGCGGGTACTCGCTCG 1380
QY 1741 CGATCCGTGACCTCGCGCATCTCTCGTCGACACCGCGCGCGCGCGCGCGCGCGCGCGCG 1800
DB 1381 CGATCCGTGACCTCGCGCATCTCTCGTCGACACCGCGCGCGCGCGCGCGCGCGCGCGCG 1440
QY 1801 GCGCACACCGTCTCTCCCGAGAGTTCACCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1860
DB 1441 GCGCACACCGTCTCTCCCGAGAGTTCACCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
QY 1861 TACACCGAGCG 1920
DB 1501 TACACCGAGCG 1560


```
Qy 1921 ATCTAG 1926
|||||
Db 1561 ATCTAG 1566

RESULT 2
AF294730
LOCUS Microbacterium sp. AL-210 linear BCT 15-NOV-2001
DEFINITION complete cds.
ACCESSION AF294730.1 GI:9885836
VERSION AF294730
KEYWORDS
SOURCE
ORGANISM Microbacterium sp. AL-210.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Micrococciaceae; Microbacteriaceae; Microbacterium.
1 (bases 1 to 1890)
Chae, J., Park, N.H., Yang, S.J. and Lee, T.H.
Molecular and enzymatic characterization of a levan
fructotransferase from Microbacterium sp. AL-210
J. Biotechnol. 91 (1), 49-61 (2001)
21413997
11522362
2 (bases 1 to 1890)
Yang, S.J. and Cha, J.
Direct Submission
JOURNAL Submitted (09-AUG-2000) Microbiology, Pusan National University, 30
Changjeon-dong, Keumjeong-ku, Pusan 609-735, Republic of Korea
Location/Qualifiers
1. 1890
/organism="Microbacterium sp. AL-210"
/strain="AL-210"
/db_xref="taxon:135249"
181..1773
/gene="lftm"
181..1773
/gene="lftm"
/codon_start=1
/transl_table=11
/product="levan fructotransferase"
/protein_id="AAG01554.1"
/db_xref="GI:9885837"
/transl_table="MTSEFSRRRTMLKTSVGLALLIGSSPTLAOSSAGLRAPAAAA
SGSLRAYHMTTPSGWLCDPQVRVTGGAYOLYVLSHGQNGPQGDHATTSDGVAFT
HGTVMPLQDFVWSGSVAVDTANTAGFCAGAVIALATQPTDGIKRYQOYLWSTD
GGTFTALPDVILNDGRTATPAEIEAENFRDPRIHMDALRGWCVIGRARTAS
FYITSLNDQWTSNFDYPNHALGGIECPDLFEMTAGDGTTRHILASMDAYGVGLPM
TYAIWTGNTGTHFIAGNLVQWLDGWDYAAVTPWKAEPETRLAIAMNWKYA
ARDVPTDASDGYNGQNSITRELRLERQEGWYLLSEPVGALGGYVTATTITPVVN
GSVLPWSGRAYEILDLSWTDAAVGLSVGRSADGARHTNICKYGSGLYVDRGPSDO
AGFALSPFTAAAPIDAGVRSVHLRVFVDTSVEVFNAGHTVLSQOVHFLPGDTGIS
LYSDGAATFSGIRIREFPTPI"
BASE COUNT 318 a 663 c 610 g 299 t
ORIGIN

Query Match 48.88; Score 977; DB 1; Length 1890;
Best Local Similarity 78.38; Pred. No. 4.5e-97;
Matches 1170; Conservative 0; Mismatches 324; Indels 0; Gaps 0;

Qy 436 GCTGTGCGCGCTCGAGCCGGGCGATCCGCTCCGGGCTCGCTCCGTCGCGCTTACACATG 495
|||||
Db 283 GCCGGCTCCGGGCGCGCGCGCGCGCATCCGATCCCTGCGCGCGCTGACCATG 342

Qy 496 ACGCCCGCCAGCGCTGCTCGGACCCCAACCGCGGTTCACACCCAGCGCGCCTAC 555
|||||
Db 343 ACCCGCGCTCGGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTAC 402

Qy 556 CAGCTACTACCTGCACTCGGACAGACACCGCGCGCGCGCGCGCGCGCGCGCGCG 615
|||||
Db 403 CAGCTACTACCTGCACTCGGACAGACACCGCGCGCGCGCGCGCGCGCGCGCGCG 462
```

```
Qy 616 ACAGCCGACGCGCTCGCTTACAGCACACGACGCGACCGTGTGCGCTCGCGCCGACATTC 675
|||||
Db 463 ACGTGCGACGCGCTCGCTTACATCATCATCGGACGCGTGTGCGCTCGAGCCGATTC 522

Qy 676 CCGTGTGGTCCGGGTGCGGCGTGTGCGGACACCGCGAACGCGGAGGTGCGGCGCGCG 735
|||||
Db 523 CCGTGTGGTCCGGGTGCGGCGTGTGCGGACACCGCGAACGCGGAGGTGCGGCGCGAGG 582

Qy 736 GCGTGTGCGGCTCGGACCGACCGACGCGCGCGTTCACGTTACCCCGCTGCGCGACCGTATCG 795
|||||
Db 583 GCGTGTGATAGCATGCGGCGACGCGACCGCGCGCGTTCACGTTACGCTGCGCGGATCGGTATCG 642

Qy 796 CTCTACTGCGTACCGCGCGGCTTACGTTACCCCGCTGCGCGACCGCGTATCGTATCGTC 855
|||||
Db 643 CTGTACTGCGTACCGCGCGGCTTACGTTACGCTGCGCGGATCGGTATCGTATCGTATCGTC 702

Qy 856 AACACCGAGGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 915
|||||
Db 703 AACACCGAGGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 762

Qy 916 GACCCCAAGATCCACTGGGACACCGCGCGGAGAAATGGTCTCGCTCATCGGAGCACTG 975
|||||
Db 763 GACCCGAGATCCACTGGGATGCGCTGCGCGGAGTGGGTGCTGCTCATCGCGCGCGCG 822

Qy 976 CGGTACCGCGGCTTCTACACCTCGCGAAACCTGCGCGACTGGACACTTGGCGCGAACTTC 1035
|||||
Db 823 CGTTACGCTCGTTCACATCGACGAAACCTCGCGACTGGCAATGGACATTCACATTC 882

Qy 1036 GACTACCGGAACACGCGCTCGCGCGGATCGAGTGCCTCGCGACTGTTTCGAGATCACCGCA 1095
|||||
Db 883 GACTACCGGAACACGCTCTCGGCGGATCGAGTGTCCGACCTGTTTCGAGATCACCGCC 942

Qy 1096 GACGACGGGACACCCACTGGTGTCTCGCGCGAGTACGCGCTACGGCATCGGCTC 1155
|||||
Db 943 GCGGACGGGACCGCGCGCTGGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1002

Qy 1156 CCCATGACGTACGCTTACTGACAGGACCTGGGACGCGCGCGCGCGCGCGCGCGCGCG 1215
|||||
Db 1003 CCGATGACGTACGCTTACTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1062

Qy 1216 CTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1275
|||||
Db 1063 CTCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122

Qy 1276 ATCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1335
|||||
Db 1123 GCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1182

Qy 1336 GCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1395
|||||
Db 1183 GCGGAGAGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1242

Qy 1396 GAGCTGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1455
|||||
Db 1243 GAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1302

Qy 1456 GCGCTGACGAACTACGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1515
|||||
Db 1303 GCTCTGGCGCGTACGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1362

Qy 1516 GCCGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1575
|||||
Db 1363 GTCGTACTGCGTGGAGTGGAGCGCGTACGAGATCGAGATCGAGATCGAGTACGATCGAGCGAT 1422

Qy 1576 GCGAGAACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1635
|||||
Db 1423 GCCGCCAACGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1482

Qy 1636 GGAAGTACGAGGACGACCTGTAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1695
|||||
Db 1483 GGAAGTACGAGTCCGACCTGTAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1542

Qy 1696 CTCGCCCGCTACTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1755
```



```

Db 1170 TCCGGCGCTTTCGAACTATGCCAATCCAGCACCACCTCCGGACCCACAGTCAACGG 1229
Qy 1512 CAGCCCGCTCTGCCATCAAGACGCGCATACAGATCAGATCGAGTCTCGCTGGGA 1571
Db 1230 CAGTTTCGACTTCCGTGGAGCGCGCGGTATGAAGTGAATCTCGATATTTCAATGGGA 1289
Qy 1572 CACCGGACGAAAGTCGCGATCTCGGTGGCGGCTCCCGCCGACGGAACCCGGCACACGAA 1631
Db 1290 CACGGCAGCAAGCTGGGAGTCTCGGTGGCGGCTCGGTCCGATGGCAGCGCCCATACGAA 1349
Qy 1632 CATCGGCAAGTACGAGCAGACCTGTAGTTCAGCGAGGAGCCCTCCGACCTCGCGGGGA 1691
Db 1350 CATCGGCAATACGCTGACGAGTGTAGTTCGATCGCGCATCTCCGGAGCAAGCGGTTA 1409
Qy 1692 CTCGCTCCCGCCCTACTCGCGAGCGCGCGCCCATCGACCCCGCGCGCATCGGTGCA 1751
Db 1410 TGGCGTGGCACCTTACACCGCGCGCGCGCCCATCGATCGAGCGCATCGGTCCA 1469
Qy 1752 CTTGGGATCTCTGTCGACACCCAGAGGTCGAGGTCTTCGTCAGCGCGCGCCACACCGT 1811
Db 1470 CTTGGGATCTCTGTCGACACCCAGAGTGTAGGTGTTCGTAATTCGGGCGCACACGGT 1529
Qy 1812 GCTCTCCAGCAGTCTCACTTCGCGAGGCGGCGACACGGGAATCTCGCTCTACACGCGG 1871
Db 1530 GGTTCGAGCAGGTGCACTTCGCGCGCGCGCGCGCCCATCGATCGAGCGCATCGGTCCA 1589
Qy 1872 CGGCGCGCACACTTCACCGGCATCTGTCGCGGAGATTCGCCAGGCGATCTAGGCGAT 1931
Db 1590 CGGTCCGCGCAACTTCACCGGCATCACCATCCGCGAGTTCGGGAACCCCATCTAAGCCTG 1649
Qy 1932 GCACACACACCGTTCACCGAAGCGCGCGCGCGCGGAGACGACGG 1975
Db 1650 CGTCCCGCGGTGGAAGGACGCGCGCGACGCTCGACGAGCGG 1693

```

RESULT 4

```

E22826
LOCUS E22826 1551 bp DNA linear PAT 18-JUN-2001
DEFINITION Levan fructotransferase gene.
ACCESSION E22826
VERSION E22826.1 GI:13024125
KEYWORDS JP 1999069978-A/2.
SOURCE Arthrobacter nicotinovorans.
ORGANISM Arthrobacter nicotinovorans
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micrococciaceae; Micrococciaceae; Arthrobacter.
Masakazu, S., Atsushi, Y. and Fusao, T.
Levan fructotransferase gene
Patent: JP 1999069978-A 2 16-MAR-1999;
NIPPON OIL CO LTD
OS Arthrobacter nicotinovorans
PN JP 1999069978-A/2
PD 16-MAR-1999
PF 28-AUG-1997 JP 1997232421
PR

```

```

PI MASAKAZU SAITO, ATSUSHI YOKOTA, FUSAO TOMITA
PC C12N15/09, C07K14/195, C12N1/21, C12N9/10//A61K38/45, (C12N15/09,
PC C12R1:06),
PC (C12N1/21, C12R1:19), (C12N9/10, C12R1:19), C12N15/00, A61K37/52,
PC (C12N15/00, C12R1:06)
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..1551 /organism='Arthrobacter nicotinovorans'.
FT Location/Qualifiers
1..1551
/organism='Arthrobacter nicotinovorans'
/db_xref='taxon:29320'

```

FEATURES

source

BASE COUNT

ORIGIN

315 a 524 c 443 g 269 t

Query Match 46.6%; Score 931.8; DB 6; Length 1551;
 Best Local Similarity 75.8%; Pred. No. 3.6e-92;
 Matches 1184; Conservative 0; Mismatches 367; Indels 12; Gaps 2;

```

Qy 361 ATGACGGCGGCATCTCAGCGCGCGCGTGTCTCAGGAGAGCGCGCGGAGACTCGCC 420
Db 1 ATGACGATGACATCTCTCGCGCGCATGCCCTGCAAGGTGAGGGGTGTGCTTGGCA 60
Qy 421 CTGATCTTCGGCGGTGTGTGCGCGCTGACGCCGGCATCGCTCGGGGTCTCGTCCGT 480
Db 61 CTTTTCATGAGCAATGCCATTTCCCGTGGCGCCCA-----CGCCAGGCAATCCCTCCGG 114
Qy 481 GCCCTTACCATATGAGCCCCCGCGGGGTGGTCTCTCGACCCCGCCCAACGCCGCTCACC 540
Db 115 GCGATCTACCATATGAGCCCCCGCTCGGGTGGCTATGTGATCGCAGGACCCGTACAT 174
Qy 541 ACCACGGCGCTACGAGCTGTACTACGTGACATCGACACAGAACAGCGCGCGCGGCGG 600
Db 175 ACAACGGCGCTACGAGCTGTACTACCTCCAGTCCGGCCAGAACAGCGACCGGGCGGA 234
Qy 601 TGGACACGCGAGCAGACCGCGGTGCGCTTCACGACACACGACCGCGTGTATGCCG 660
Db 235 TGGACACGCGACACCGCGGAGGTGCTTTACACCCACCATGAGTGTGATGCCA 294
Qy 661 CTGCGGCGCGATTTCCCGGTGTGTGCGGGTGGCGGTCTGTCGACCGCGGAAACACGCA 720
Db 295 ATGCAACCGGACTTCCCGGTGTGTGCGGATCGGACGTAGTGGACACCGCCACACCGCC 354
Qy 721 GGGTTCGGCGCGCGGGTGTGCGGCTCGGACCCAGCGCGGAGCGGCGTCCGCAAG 780
Db 355 GGGTTCGGCGCGCGGAGTATCGCGCTCGCCACCAACCCACCGAG-----GAAA 408
Qy 781 TACCAGGAGCAGTACTCTACTGTGACGACGCGGGGTTCAGGTTACCGCGCTCGCC 840
Db 409 TTCAGGACAGTACTCTTACTGTTCACGGATGGCGGTACTCTTACCGCATTCCT 468
Qy 841 GACCCCGTCACTGTCACACCGCGGTGCGCGCGCGCGCGCGCGCGCGCGGAGATCGAAG 900
Db 469 GACCGGTTCATTGTGAACACTGATGGACGAGCGCGCGCGCGCGCGCGCGGAGTGGAA 528
Qy 901 GCGAGTGGTTCGCGGACCCCAAGATCCACTGGGACACCGCGCGCGCGGAGATGGGTCTG 960
Db 529 GCAGAAATGGTTCCCGGACCGCGGAAATTCATGGGACGCGCGCGCGCGCGGATGGGTCT 588
Qy 961 GTCATCGGACACTGCGGTACGCGCGGTCTTACACTTCGCGGAACCTGCGGACTGGACA 1020
Db 589 GTCATCGGAGGCGCGGTACGCTGCGCTTCTACACTTCTCCCAACCTGCGGGATTGCAA 648
Qy 1021 CTTGCGCGCAACTTCGACTACCCGAAACACCGCGCTTCGCGGCGCATCGAGTCCCGACTG 1080
Db 649 TGAAGTCCAACTTCGACTACCCCAACACCGCGCTTCGCGGCGTATCGAATGCGCGGATCT 708
Qy 1081 TTCGAGATCACCGAGAGGAGACCGCACTGGGTGTCTCGCGCGCGCGCGCGCGCGCGG 1140
Db 709 TTCGAAATGACCGAGGAGCGGAAACCGCGCACTGGGTGTTCGGGCGGAGATGGAGCGG 768
Qy 1141 TACGGCATCGGCGCTCCCATGACCTAGCGCTTACTGGACAGGCACTTGGAGCGCGAGCAG 1200
Db 769 TACAGCATCGGCTTGGCCATGACCTTTGCGTACTGGACAGGTTTATGGAAGGCGACAGCA 828
Qy 1201 TTCACGCGCGAGCACTTACCGCGCAATGGCTCGACTTGGGCGTGGAGTGTGTAGCGGCG 1260
Db 829 TTCATCGCGGACAACCTCACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 888
Qy 1261 GTCACCTGGCGCATCGATCGAGCGCGCGCGCGAGCAAGCGCTTCGCCATCGGTGTGATGA 1320
Db 889 GTGACCTGGCGCGCGGTGGAGACCTTGAGACCAAGCGGTTCGACAGCGGTGTGATGAAC 948
Qy 1321 AACTGGAAGTACCGCGCGCGAGCTCCCGACCGACCGCATCCGAGCGGTACAACGGCGAG 1380
Db 949 AACTGGAAATATGCGCGCGCGCAACGTCGCCACGCGCGCTCCCGATATGCTATAGCGGCAA 1008

```

```
QY 1381 AACTCGATCGTCGGGAGCTGGGCTCGCCGACACCCCTGGCGGCTGTGTACACCTCTCTG 1440
Db 1009 AATTCATCACCGAGCTAGGCTCGAGCGCCAAATCGGGGGTGTGTATACCTTGTC 1068
QY 1441 AGCACCCCGTGGCAGCGCTGACGAACATACGTACCGCCACACACACACTCCCGACGG 1500
Db 1069 ACACGCCCGTTCGGCGGCTTTCGAACATATGCCACTTCAGACACCACTCCCGACCGC 1128
QY 1501 ACCGTGACGAGCGCCCTCTGTCATGGAACGACGCGCATAGAGATCGAGTTCGAC 1560
Db 1129 ACAGTCAACGGCAGTTTCGTACTTCGGTGAGCGCGCGCGTATGAACATGCACTCGAT 1188
QY 1561 ATCGCTGGGACACCGCAGAACGTGCGCATCTCGTGGGCGGCTCCCGCCGACGAAAC 1620
Db 1189 ATTTATGGGACACGCGAACGTGGGAGTCTCGGTGGGCGGCTCGTCCGATGGCAGC 1248
QY 1621 CGGCACACCAATCGGCAAGTACGAGACAGACCTGTAGTCGACGCGGAGGACCTCCGAC 1680
Db 1249 CGCATACGAACATCGGCAAAATACGGTGACGAGTTGTAGTCGATCGGCGCATCTCGGAG 1308
QY 1681 CTCGCGGGTACTCGTCCGCCCTACTCGCGAGCGCGCGCCCATCGACCCCGCGCC 1740
Db 1309 CAAAGCGGTATGCGTGCACCCCTACACCGCGCGCGCGCCCATCGATCGGAACGCC 1368
QY 1741 CGATCGGTGACCTCGGATCTCTGTCGACACCCAGAGGTCGAGGTCTTCTCAAGCC 1800
Db 1369 AGATCGGTGACCTCGGATCTCTTGTAGACACCCAAAGTGTGAGGTGTCTGTAATTC 1428
QY 1801 GGCACACCGTCTCTCCAGCAGTCCACTTCGCGAGGCGGACACGCGGAATCTCGTTC 1860
Db 1429 GGGACACGCTGTTTCGACGAGTGCATCTCGCGCGCGGCGGACACGGGATCTCCCTC 1488
QY 1861 TACACGAGCGGCGCCCGCACATTCACCGGCATCTGTCGCGGAGATGCGCAGCGC 1920
Db 1489 TATCGGACGCGGCTCGCGCCAACTTCACCGCATCACCATCGCGAGTTCGCGGAACCC 1548
QY 1921 ATC 1923
Db 1549 ATC 1551

RESULT 5
E22825
LOCUS E22825 1452 bp DNA linear PAT 18-JUN-2001
DEFINITION Levan fructotransferase gene.
ACCESSION E22825
VERSION E22825.1 GI:13024124
KEYWORDS JP 199069978-A/1.
SOURCE Arthrobacter nicotinovorans.
ORGANISM Arthrobacter nicotinovorans
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micrococciaceae; Micrococccaceae; Arthrobacter.
REFERENCE
AUTHORS Masakazu, S., Atsushi, Y. and Fusao, T.
TITLE Levan fructotransferase gene
JOURNAL Patent: JP 199069978-A 1 16-MAR-1999;
NIPPON OIL CO LTD
OS Arthrobacter nicotinovorans
PN JP 199069978-A/1
PD 16-MAR-1999
PF 28-AUG-1997 JP 1997232421
PR
PI MASAKAZU SAITO, ATSUSHI YOKOTA, FUSAO TOMITA
PC C12N15/09, C07K14/195, C12N1/21, C12N9/10//A61K38/45, (C12N15/09,
PC C12R1:06),
PC (C12N1/21, C12R1:19), (C12N9/10, C12R1:19), C12N15/00, A61K37/52,
PC (C12N15/00, C12R1:06)
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1. .1452
FT /organism='Arthrobacter nicotinovorans'
FT Location/Qualifiers
```

```
source 1. .1452
/organism='Arthrobacter nicotinovorans'
/db_xref='taxon:29320'
BASE COUNT 299 a 491 c 416 g 246 t
ORIGIN
Query Match 45.5%; Score 909.6; DB 6; Length 1452;
Best Local Similarity 77.0%; Pred. No. 9.3e-90;
Matches 1123; Conservative 0; Mismatches 329; Indels 6; Gaps 1;
QY 466 CGGGGCTCGCTCCGTCGCTTACCACATGACGCCGCCAGCGGCTGGCTCTGCGACCCC 525
Db 1 CAGGCGATCCCTCCGGGCGATCTACCACATGACGCCGCCCTCGGGCTGGCTATGTATCG 60
QY 526 CAAACCGGGTCAACACCCACCGCCCTACACAGCTGTACTACTGCTACCTCCGACGAGA 585
Db 61 CAGCGACCGGTACATACAAACGCGCTTACAGCTCTACTCTCCACTCCGCGCCAGAAC 120
QY 586 AACGCGCCCGGCTGGGACACCGGAGCAGCAGCGGCGTGCCTTACAGCAACAC 645
Db 121 AACGCGCCCGGATGGGACACCGGAGCAGCTTCCCGCTGTCGGATCGGCACTAGTGGAC 180
QY 646 GGCACCGTGTATCCGCTCGCGCCCGACTTCCCGCTGTGTCCGGTTCGCGCTGCTCGC 705
Db 181 GGAGTGGTGTATCCCAATGCAACCGGACTTCCCGCTGTCGGATCGGCACTAGTGGAC 240
QY 706 ACCGCGAACACGCGAGGTTTCGCGCGCGCGCTGTCGCGCTCGCGACCCAGCGGAC 765
Db 241 ACCGCGAACACCGCGGCTTCGCGCGCGCGCAGTATCGCGCTCGCCACCCACCCACC 300
QY 766 GAGCGCTCCGCAAGTACCAGGAGCAGTACTCTACTGTCGACCGAGCGGCTGTCAG 825
Db 301 GAGC-----GAAATTCAGGAACAGTACTTACTGTCGACGATGGCGGCTACTCC 354
QY 826 TTCACCGCTTCGCGACCCCGTCACTGTCACACGAGCTCGCGCGCGCCACACCGCC 885
Db 355 TTCACCGCATTCGCTGACCGGCTCACTGTGAACACTGATGACGAGCGGCGCACCCGCC 414
QY 886 GCCGAGATCGAGAACCGCGAGTGGTTCCGCGACCCCAAGATCCACTGGGACACCGCCGC 945
Db 415 GCGGAGGTGGAGAACGCAAGTGTTCGCGACCGGAAAAATTCACTGGGACGCGACGCG 474
QY 946 GAGAAATGGTCTGCTCATCGGACGACTCGGTAGCGCGGCTTACACTCTCGCGGAC 1005
Db 475 AACGAGTGGTCTGTGTATCGCGAGGCGCCGCTACGCTGCTTCTACACTCTCTCCCAAC 534
QY 1006 CTCGCGGACTGGACACTTCGCGCAACTTCGACTACCGGAACACCGCCCTCGCGGCAATC 1065
Db 535 CTGCGGGATTGGCAATGGAGTCCAACTTCGACTACCCCAACACCGCCCTCGCGGATATC 594
QY 1066 GAGTCCCGGACCTGTTTCGAGATCACCGAGACGAGGAGACGCGCACTGGGTGCTCGCC 1125
Db 595 GAATGCCCGGATCTGTTGAAATGACCGAGGAGACGGAACCCGCGACTGGGTGTTCCGG 654
QY 1126 GCCAGATGGAGCGCTACGGCATCGGCTCCCGCATGAGTACGCTACTGGACAGCGACC 1185
Db 655 GCGAGCATGGAGCGCTACAGCATCGGCTTGGCCCATGACCTTTGCTACTGGACAGTTCA 714
QY 1186 TGGGACCGGAGCAGTTCACCGCGACGACCTCACCCCGCAATGGCTCGACTGGGCTGG 1245
Db 715 TGGACGCGCACAGCATTCATCGCCGACAACTCACACACAGTGGCTTGACTGGGATGG 774
QY 1246 GACTGTTACGGGCGCGTCACTGGCCATCGATGACGCGCGCGGACGACCAAGCGCTCGCC 1305
Db 775 GACTGTTACGGGCGCGTCACTGGCCGCGGCTGGGAGGACCTTGAGCAAGCGGCTTCCG 834
QY 1306 ATCGCGTGAACAACTGGAAGTACCGCGACGAGTCCCGCCACCGACGACATCCGAC 1365
Db 835 ACAGCGTGGATGAACAACACTGGAAATATGCCGCCCGCAACGCTGCCACGAGCGCTCCGAT 894
QY 1366 GGCTACAAACGGCAGAACTCGATCTCGCGGAGTTCGCGCTCGCCGCGACAGCTCGCGCG 1425
Db 895 GCGTATACGGGCAAAATTCATCACGCGCGGAGCTCAGGCTCGAGGCGCAATCGCGCGC 954
```


Db 682 GAGACAGCGGACGCGTGGGTGCGCCCTCATCGCGAGCGGACCGGGTTTGTCTTAC 741
QY 994 ACTCTCCGCAACTCGCGGACTGGACACTTCGCGCAACTTCGACTACCCGAAACACGCG 1053
Db 742 ACTCTCCGCAACTCGCGGACTGGAC-----GCGGACGGGAGTACGTCAACGCGCG 795
QY 1054 CTCGGCGGATCGAGTGGCCCCGACCTTGTGAGATCACCGCAGACGGGACAGCCAC 1113
Db 796 ATCGGCACCATCGAGTGGCCGACCTTCTCCGCTTCGGGCGGACGAGCGGACGACCCAC 855
QY 1114 TGGGTCTCGCGCCAGCATGAGCGCTACGCGATCGGCTCCCATGATGAGTACGCTAC 1173
Db 856 TGGGTCTCGCGCCAGCATGAGCGCTACGCGATCGGCTCCCATGATGAGTACGCTAC 915
QY 1174 TGGACAGGACCTTGGGACGGCGGAGTTCACCGCGGAGCTTCCACCGCGCAATGGCTC 1233
Db 916 TGGACGGGCTCGTTCGACGGGTTCGACCTTCGTGCGGAGTTCGTTCGCGGAGTGGCTC 975
QY 1234 GACTGGGGTGGGACTGGTACGCGGCGGTGACCTGATCGGCCATGATCGACGCGCGAGACC 1293
Db 976 GACCACGGGTTGACTGGTACGCGGCGGTGACCTGGGAGGACCCCGCGACCGCTCGAG 1035
QY 1294 AAGCGCTCGCATCGCTGGATGACAACTGGAAGTACGCGGACGCGGACGCTCCCGAC 1353
Db 1036 CGCGCGCTCGCGATCGCTTGGATGAACAACCTGGGACTA---CGCGACGGGCGCGGAC 1092
QY 1354 GACGATCCGACGGCTACAAACGGGAGAACTCGATCGTCCGCGAGTGGGCTCGCGGCA 1413
Db 1093 TGGCGCGGACGGATTACCGCGGACCGGACTCGATCACGCGGAGATCGCGCTCGCGCG- 1151
QY 1414 GAGCTGGCGGTGGTACACCTCTGTAGACACCCCGTGGGAGCGCTGACGAACTACGTC 1473
Db 1152 --TGCGGTGCGGGTACAGCTCTCTCTCGCTCGCGTCCGCTCGAGCGGCTCCAGGAC 1209
QY 1474 ACGCGCACACACACTCCCGACCGGACCTCGAGGAGCGGCGCTCTGCGCATGGAC 1533
Db 1210 ACGCACACACGCGGCTCGGCGAGCTCGCGCTCGAGCGGCTTACCGAGTTCGCTACGCG 1269
QY 1534 GGACGCGATACGATCGAGTCTGACATCGCTGGGACACCGGACGACGCTCGGCACTC 1593
Db 1270 GCGACGCGTACGATACGACGCGGATCTCGCGGACGAGGACGACGACGCTCGGGTTC 1329
QY 1594 TCGGTGGCGGCTCCCGACGAGAACCGGACACGATCGGACATCGGACGAGTACGAGAGAC 1653
Db 1330 CAGCTCGCGGCTCGCGGACGAGCGGACGCGGACGCGGCTCACCAGCGGACTAC 1389
QY 1654 CTGTAGTCCGACGAGGACCTTCGACGCTCGCGGCTGCTGCTCGCGGCTTACTCGGA 1713
Db 1390 GCTTACTTACCGGCGGACGAGCGGCGGCGGCGGCTGCTGGAAGG-----TCGAGA 1441
QY 1714 GCGCGCGGCGGATCGACCGGCGGCGGCGGATCGCTGCACTCGGATCGCTCGGACAC 1773
Db 1442 GCGCGACGCGGCTGGA-----CGACCGGACACCGTGGAGCTCCGATCTCTGTCGACGCG 1497
QY 1774 CAGAGGTGAGGTCTTGTAAACGCGGCGGACGCGGCTCTCCAGGAGGTCCACTTC 1833
Db 1498 ACAGCATGTCAGGTCTTGTCTGCGGACGCGGCGGCTTACACGACTCAGCGGAGGTTCGCG 1557
QY 1834 GCGGAGGCGGACGCGGATCTGCTTACACCGGCGGCGGCGGCGGCTTACCGGCG 1893
Db 1558 CCGTGGGGGACAGAGGCTCGCTTACACTTGGGCGGCGGCTTGGGTGTTCGCGAT 1617
QY 1894 ATCTGCTGCGGAGATTGGCAGGCGATCTAGGGGATGACACACACGCTACCGGAA 1953
Db 1618 CTGCGGATCACCGAGTTGCGACACCTCGCGGACGTCGCGGCGGCTCTCTCGCGGACTTCG 1677
QY 1954 GCGGCGGCGGAGGACGAGG 1975
Db 1678 AGGCGGACAGTGGGCGGCGG 1699

RESULT 7
AX211705

LOCUS AX211705 65140 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 1 from Patent WO0159126.
ACCESSION AX211705
VERSION AX211705.1 GI:15523937
SOURCE Streptomyces noursei.
ORGANISM Streptomyces noursei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 65140)
AUTHORS Zotchev,B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
Gulliksen,O.M.
TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 1 16-AUG-2001;
Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
(NO)
FEATURES
source Location/Qualifiers
1..65140
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"
BASE COUNT 8270 a 25171 c 22273 g 9426 t
ORIGIN
Query Match 7.4%; Score 147.6; DB 6; Length 65140;
Best Local Similarity 45.4%; Pred. No. 5.7e-08;
Matches 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;
QY 8 ACCCGACTTCCCTCGAGCACCGTCCCTACCGCGGCGGCGGCGGCGGCGGCGGCGG 67
Db 19803 ACGGCGGAGACACGACCGCGCGGCGGCTTCGGACTTCACCGCGGCGGCGGCGGCGG 19862
QY 68 TCCTCAGCTAGACGCGGCGGCGGCTTCCTCGAGGTCTTCGCGGAGCGGTGAGGCGACTTGC 127
Db 19863 AACACGCGCGGCTTACGCGGACCTCGGCGGATCAGCGCGGCGGCGGCGGCTTGCCT 19922
QY 128 CGAACCTGTCTCTTGGGGGCGGCGGCTTACGCGGAGCTTCGAGAGCGGACGCGGCGAG 187
Db 19923 GGAAGAGGCTCTCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 19982
QY 188 GAACCTGTGACGTCGCGGATCGAGTCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCG 247
Db 19983 CCGGCGGAGGACACGCTACCATCGCGGTCTACGAGCGGCGGCGGCGGCGGCGGCGG 20042
QY 248 CTGCGCGGCTTCTGGGCTGACGAGCGCTCCACCGCGGACGCTTCCTTCTACCGGTCG 307
Db 20043 TCGACTCCCTGTCTCCCGCGAGGTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20102
QY 308 CGAACGAGGCTGAGCTTCGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 367
Db 20103 ACCGGGACTCCCTCTTCCAGTTCGAGTGGACCGCGGCTCCAGGCGGCGGCGGCGGCG 20162
QY 368 CGGCGATCTACGCGCGGCGGCTTCCAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 427
Db 20163 CGGCGA-----CCGTCGCGGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 20213
QY 428 TCGGCGGTGCTGTGCGGCTTCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 487
Db 20214 TCCGCG-----CCACCGGATCCGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGG 20264
QY 488 ACCATGACGCGCGCGGCTGGCTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 547
Db 20265 CCGAGCGCGGAGGCGGCGGCTCCCGGCGGCTTCACCGGCGGCGGCGGCGGCGGCGG 20321
QY 548 GCGCGTACGCTGTACTTACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 607
Db 20322 GCGCGCGGCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20381


```

/codon_start=1
/transl_table=11
/product="Nysg"
/protein_id="AAF71763.1"
/db_xref="GI:8050837"
/translation="MASPDDLEERTAPRARRLGLLRPHRRSVALAVSMGVGGIVL
NAFGPLLLGRVTDLIADGVLGVPAGPDIIDFALGRLLLVLLALVYVASLFLMAQGR
LVASAWRTIHELRRDAKEKRLRPLRDPAGPAGELLSRTINDINLQOQLQTLAE
LITFISLLMLVLLISPSLAVVLLSVFSAIARAKISKRAQPHYAAQWMSANGTL
NAHVEVCTGHALLKGFDRRAAEERFACNDVAVRAAKAQFASGAPMEVMEFVANL
GVLLVAVGAKNPLNGTLTGDOVAFILYARQFSQPIVEIASVAGRLQSGTASQORVF
TYLDAPQAPKPLRPGTPARAEGVEFTDNFSRYSPDPLIENLSLVEPSTVAIVG
PTGAKTTLGNLWRFEPDPSGRILLQDGTDTATRRDLRFRGLQDQTLWLFQGTIA
ENIAYGAPGACADATIEAARATCADRFRTIPLQGVDTVLDDESGTSGAGERKOLLTVAR
AFLLPAPVLDIADTSVDTREVLIQRAMSLRAGRTSFVIAHRLSTIRDAIDLIVM
DAGRIVEQTHQDLCQAGLYARLHAARTHTPTAGAAAG"
complement(2662..4416)
/gene="nysg"
complement(2662..4416)
/gene="nysg"
/function="presumably involved in efflux of nystatin"
/notes="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="Nysg"
/protein_id="AAF71764.1"
/db_xref="GI:8050838"
/translation="MLRLRLRAQLPYAWATAALVALQVLQVILGTLTLLTGLAALIDQ
GVNGDGRITELGVGMVVALVOITAAALGAALAAARTATAMGRDLRSALFRILDFS
AREIGRFTCSLLTRSVNDVOQVONLAOTGFCIVVCAPLMCLGSLVLLARODVPLALL
LVALVYVAVCFGLLLARMGTLYARMQTLDRGLRLREALITGVRVRSFRVDRHERA
RPAQNVAVVRSRGRILAIATMLPVLVILLMNGFTVALLWTGSHRIDAGRMPIGLSLA
ELCAYLITMSVWMLAFVFLVPRVPRCAGRIAEVLDTCSSVAPAPVPRGPGAGRI
BLSYAGVPAEFPFLVDVLTVEPGERIAVLGTSKGTLLNLVRLRLADATAGAV
RVGGTDRVLAALTAAGGVGPORPYLFSGTVASNLFRGRPDATDEBELVALVAQA
ADFARMPDGDAETIAAGGVNSGORLRLSARALLRFRPEILYFDGCFSDALDQATDA
ALRYALPYTAGATVITVAQRISAGRDADRIVLDGRVVAQGTGHDVLLRSPYREI
ALSQTEEAHAGLAGRP"
4714..5748
/gene="nysdiii"
4714..5748
/gene="nysdiii"
/function="presumably involved in mycosamine biosynthesis"
/notes="putative dcbp-mannose-4,6-dehydratase"
/codon_start=1
/transl_table=11
/product="Nysdiii"
/protein_id="AAF71765.1"
/db_xref="GI:8050839"
/translation="MSKRALITGTQDGSYLABHLLSQGVQVWGLINGQANPKRSRV
SRLASELDFIDGDLMDQSLVSADVTQPDVEYNLGAISFYPMNQQAELVTEYNGMG
VLRMLEAIRMYSGLSRTSRVSPRGQIRFYQASSEMFGKAAETPQRETTLPHPSPYG
AAKAYGHYITRNYRESFCOMYAVSGMLFNHSPRGQEPVTRKISLAVARIKQGLQDL
ALGNLDVDRMGYAGDYVRAMHLMQDQAGDDYVIGTQMHSVSRDARIAREHVLGNW
EDYVITDPLVRPAEVEVLCADSARQRLGWKPDVDFPTLIRMVWSDLAQVSRNGQ
YGDVLLAANW"
5930..34363
/gene="nysi"
5930..34363
/gene="nysi"
/function="responsible for condensation steps 9 to 14 in
the nystatin polyketide backbone synthesis"
/notes="polyketide synthase"
/codon_start=1
/transl_table=11
/product="Nysi"
/protein_id="AAF71766.1"
/db_xref="GI:8050840"
/translation="MDNEQKLRDYLKLATADLRRLRRRHHKLESAAQEPVIAIMTCR
YPGVRSPEDLWRMVEAGEHCVTPPTDRGMDLEALAAAPASGGLFHADPDFDADF
GISPREAVMPQORVVLVESAWAEFERAGIDPTSVKGSRTGVEIGAMAQDYRVGPDAG
AGEPQLGTMGVSLSGSLSTFTGVPANVTDTACSSSLVAVHLATQALRAGECTLAL
AGGVITMSGPTFTIEMRGQGLSADGRCSRSPEDTADGTGWAEGVGLVLERLSDAVRN

```

```

CHEILAVYRGTAIVNQDGASNGLTAPNGPSQOQVITQOALVYNARLAAGDDIDVVEAHGTGT
TLGDPVEAQAALLATYQGNRPADRLPLLLSGVSNLSHSTQAAAAC?AGVIRKMWAMRHGTL
PRTLHAEPTTHVDMVSGQAVRLLTDTTDMPTAPRRAAVSVSGISGTSNAHTIIEQAP
EPQPEDAATAQDAAAGSTPATAPVVPVGVLLSGRTPD...3QAAALRAALDTRGRP
DLDLHSLATTTRAGFEHRAVLTATLHPALDGLTADADAPAAAPAWITGTTTRAET
RLAVLTISQGAQRLGAGRLAARPAFATAALDAALDAPTHLDRPLREVMGCTDAALL
DRYAYAPALFAVEVALYRLIESGVRPHDLHAGSDVGAIVAAHLSAGVLSLADAATVA
ARGRMQALPDGGAMIAVOASEADVAPLLAGHEDQVAIAVNGVAVLVLSGAEATVTA
LAEQLAADGRKTRRLRVSHAFSPHSLPEMLDAFRVAVLDTLQPLLVNLTGKPA
TVAQLAADGRKTRRLRVSHAFSPHSLPEMLDAFRVAVLDTLQPLLVNLTGKPA
DAVTLPALRACRPEHTLTLAGLHVHGTALDWTGCFAGTCARTDLTPYAFORRY
WPKALQSGTADLRSVGLCAAHPLLSAAVSLADAGGTLTGLRSLQTHPLMDHTRVG
TTLPGLTAFLEAVRAGDEVDREELTAAPLLPEQGGVQVQVLMIGNPDVSRRRT
VNIHARPDTDDTPTAHATGTLTTADASRLQPASSEQGGTLAGDPHAPALDAQMP
AGAPFLPDHGYDLRADGGFGYVPGQLRAWRGGDYVYAEVLEPAGRSDAEAFGL
HPALLDAALHAAPTGLGERGGPPLFSWEGVSLHAGGATTLRVLTATPVDALATLV
HALLGAPVLSVDSLVSATQOLDTAAAVARDALFLDWTTPVOTPTDPPGVALLGA
DPFGLLTHAGFADAPAYPDIAAANAAGDPVTTVLSLACTGDDAADPARSAHRCAAE
ALAAVQTMWLDHHERFAARLVFTRGATVGRDVAANAAGVGLVRSQSNPCGCFALVDL
DPDGAAGAAALVAVLVSQGLAVRDLVRLVRRPLTEVAGADGTGDBGDGGSG
VSFSGEAGVLTGGTGLGAVLRAHLVAEYGVRLDLSVRSRGERAGAGELVAELAGV
GARVYVACDVTDRAAVVELVGHAVSAVHAAGVLDGDMVGLTGERLSAVLRPKVD
AVHLHEATRGILDLDAAVVFSSLAGVFCGPGOANYAANAFDLALMTRRRABGLPGLS
LAWGPWEGSGMTGTLTDVDAERLARSGVPLSVAQGLALDAAVAGTATCVPVRLD
LPVLRARGEPPLLRSLLRARRARAAGSITAGNLAQRLRLRLEDGDEDMVLDLVRG
QVALVLGHATGDDVAGRAFRDLGDFDSLVAELNRNLMTVTLGLRPLTLVDFPYVRH
LATVLDLGLTDAEATVQAAVADDDPIVIGMACRYPGVSSPDLNRLVTEGT
DAVSGFTPNRNGWESLYHPDPHGTSTYRSGFLHEAGBDFPGFQMSPREALATD
SOORLLPNSWEATERAGIDPVLSRSGRTYFAGVMYSDYSAMLASPEFEGFEQALSD
PSLASGRVATYLGLEGPAVTDTCSSLVAMHMAQALRSCECLLAGAGVYVMSSTP
AVYDFDQRGLSPDGRCKAFADAADGVNSEGVLVLRQSDAVRNGHELAVVRG
SAVVDQASNLTPNGPSQOQVITQOALVYNARLAAGDDIDVVEAHGTGTGTPTEAQA
LLATYGRDEREPPLLLSGVSNLSHSTQAAAAC?AGVIRKMWAMRHGVTPLTHVDA
SHVMSSEGAELTSEQAAWPTETRVRRAGSVSGISGTSNAHTIIEQAPKAAIAAGA
ARRTPAPVPLVLSGRSALRQAALHGLHQAQDAELVDLVSALTTRSRFEOARAA
VVAODRQILASIGALAARDPDAVVEGEAARGRTAVLFTGQSSORAAAMRELHEVO
PEFAAFDVCACVFPDLDRLRLEVVFADSGSDEAALLDETQWTPALFAVEVALFRL
VEDESVRPFVAGHSIGETIAAHVAGVLTLEDAKLAARATLQKLTGAMTIAQA
TEDEIAAHLDDTVAIAAVNGPSQVSIQDEAAETIAATFAERGGKTLRLRSHAFHS
PRMDGLDAGATFIVAEGTYRAPRIPLVSDLTGRDADAEVCTAEYVVRVREAVFAD
CVRTLRODAGATFLELSDGLLTAAEDTLGDHDAELVPMLRAGAELEAAATALAR
LVQGVGDVAAVYLAGTCARRDLPYAFQHAYWYPOLTPAALAAADPADQOOLWAA
LQGVGDRELADIILGEGDLPDLSLLPALTSWRGNQEKHLIDTLYRVEVTRLSKP
TAPVLDGTWLLVASDATAADQPALDGLDALGSHGARVRLLLDDSCADRVALERL
ADTADVDAATQVLSVLPDLDERDADDPLTGLALTVALQALDADTAGQRLWTATRG
AVSTNPADVTTPVQAAAGLGRGVLEHPRLWGLVDLPQVDFERAGQRLAGILAVK
DAPDGEDVALRATGVSGRVLRHTVEALPTAAEFTATGTLTGCTGLGAEARWL
ARAGQHLVLTSSRRGPADGAELRAELGEGVPSVSVACDVADRDLAAVLTALPEE
LPLTGVTHTAGVGHYGLDITLSTAEFAGLTAAKLAGAHLDAALLADRELDFVFGSI
AGVWMSGNQSAAGNAYLDALALHRRRAGLAATSVMGPWAAAGMAADDVASETLRR
OGLGLDLPAPAMTELRRVAVDVTVTVDVQRYAPLFTSARPSALIAGLPEYRAL
ADERTEDQATGASEVTVRRAALAEPEQLRLTDLVTPESATV.LGHSSADAVPEGRA
RDVGFDSLTAVELKRKLGATGLSLPTMVFDPYTPLEAQLRAELIICAVLEVAGPV
ATGADDEPIAIIIGMACRFPGGVSPPELWDLVAGSDTAISEFPVNRGQWOTGHLDPD
PDRGCTTSTOGGELHEADEFPITFFGSPREALVMDQQRLLLTETTESPERAGIRP
ETLRTSTTGTGSSYQYGLGAGDTGEGHMTGSSPSVLSRLSLVFLGEGPVTVD
TACSSLVALLHACQLRNGESNAVAGGATIMTTPNPFIASFQRLAKGCRKAFS
DDADGMLTAGGVVVLVERLSNDAQRNGHPVLAVRGSAINQDGSALATPANGSQOR
VIROLANARLPDIDALEHGTGTPLGDPLEAQLFATGDRDRPESALLLSGVKS
NIGHTQSAAGIASVIKWYMARHSELPTLHADAPSHVDSACTVRLTTOARAWPET
GRPRRAVSSFGISGTNHNVLLEQAPVADTFAERPAVAPVPIAGVYVWVVTARSAA
ALRQOAKERLHAETVGTALPAAGPLDGLSLVSARFAERHRAVVVPAGDIDLAALR
AVATDGPSPVAVGVADVGEKTVFVPGQSGQWGMGSQLDESDEFAERABECAAAL
AEFTDWSLVDLVRGVVAPSLERVVQPSAFVAVVSLAALWRSGVLPDVAVHSGQ
EIAAAVYSGALSRLDGRVAVLRQAIGRALAGRGMMMSVALSVLDVLEPRLVEPEGRV
SVAVNGPSPVAVAGEPELDAHLARLTADDIARRIADVASHSHOVEDLHEELLEV
LAEIAPRTSEVFFSTVTGDMWLDTARMAGTFRNLRGRVFAADAVOLLAAEYRAFV
EVSSHPVSMVAVSAIDGAPVAVAGTFRDQGGTFRLLSAAEFVVRGVVDVWAGL
FEGTGSARIDLPTTYAFQHEHLWVPPAPVAAADPDADFTAVEDGDSVALLTAAG
TDEDSVAALPALTSWRRARRDRSTDAWRYRVAWKPLGGTLPHPLFTGTLTADG
IDDTDVAGALTGYAEVRLVDEEDVAVRLRAGAEVDTGTVLSVLAARERTDAG
PDTLSLGTALTGYAEVRLVDEEDVAVRLRAGAEVDTGTVLSVLAARERTDAG
AALEHPQWGGTLDLPAALDARAQRLLAAVLSGALGSDQLAIRPSGVFTTRIVRAEA

```


Query Match 7.4%; Score 147.6; DB 1; Length 123580;
Best Local Similarity 45.4%; Pred. No. 4.6e-08;
Matches 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;

Qy 8 ACCCCGACTTCCCTTCGACGACACACGTCCTCCCTACCGGCGGACGCGCCGCGGACTGCG 67
Db 79656 ACCGCGGGGACACGACGACGCGCGGCTTCGAGCTGCACCCGCGCTGCTGGACGCGGCAC 79715

Qy 68 TCCTCAGCTAGAGCGGCGCTCTCGAGGTCCTCGTGGGAGGCTGAGCGACTCGGT 127
Db 79716 AACAGCGCGCGCTACGCGGACCTCGGCGCATACGCGGCGGCGCTGCGGTTCGCGCT 79775

Qy 128 CGAACCTTGGTCTCTCGGGGCGGCGGTGTGACCGAGCTCGAGCGGACGCGGCGGACG 187
Db 79776 GGAAGGCGTCTCGCTCGCGCGCGCGGCGCACACCGTCCGCGCGCGGATCGCCCGG 79835

Qy 188 GAACCGTGCAGTACGCGATGACGTGAGGCGCGGAGGATGTGACGCGCCCTGAAC 247
Db 79836 CCGCGGAGGACACCGTACCATCGCGCTTACGACGCGCGCGGCGGACCGTGTGTCCG 79895

Qy 248 CTGCGCGCGTGTGAGGAGCGGTCCACACCGGACAGCTCTCTTCTTACCGCTGCC 307
Db 79896 TCGACTCCCTGCTTCCGCGAGTCCCGCGGAGGACACCGCGCGCGCGCGGCGGCGGCGG 79955

Qy 308 CGAACAGGCTGAGCGTCTCGTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 367
Db 79956 ACCGCGACTCCCTTCTCACGCTGAGTGGACCGCGCTCCAGGCGCGCGGCGGCGGCGGCGG 80015

Qy 368 CGGCGATCTACGCGCGCGGCGTGTCTCAGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 427
Db 80016 CGGCGA-----CGTCCGCGTCTCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80066

Qy 428 TCGCGGTGTGTGCGCGCTGCACGCGCGGCGATCCGCTCGGCGGCTCGCTCGCGGCT 487
Db 80067 TCCGCG-----CCACCGGATCGGACACCGCGCGCGCGGCGGCGGCGGCGGCGGCGG 80117

Qy 488 ACCACATGACGCGCGCGCGGCTGCTCTGCGACCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGG 547
Db 80118 CCGACGCGAAGGCGCGTCCCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 80174

Qy 548 GCGCTTACCAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 607
Db 80175 GCGCGCGCGTCCCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80234

Qy 608 ACAGGACGACGACGCGGCGTCCCTTACGACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 667
Db 80235 AGTGGCTCGGAGACGCGTTCGCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80294

Qy 668 CCGACTTCCCGGTGTGCTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 727
Db 80295 CCGACGCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80354

Qy 728 GCGCGCGGCGGCTGCTGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 787
Db 80355 AGAACCGCGCGGTTTCGCGCTTCTGACCTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80414

Qy 788 AGCAGTACCTTCTACTGTGCGACGCGGCGGCTTACGCTTACCGCGGCTGCGCGGCGGCGGCGGCGG 847
Db 80415 AGACCTTGGCGCGGCTTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80471

Qy 848 TCATGTGAACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 907
Db 80472 GCACCGAGGTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80531

Qy 908 GGTTCGCGG 967
Db 80532 GGAACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80591

Qy 968 GACGACTGCGGTAGCGGCGGCTTCTACACTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1027
Db 80592 CCGCGACCTGCTGCCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80651

Qy 1028 GCAACTTCGACTACTCCGAAACACACGCTCGGCGGATCGAGTGCCTCGGACCTGTTTCGAGA 1087
Db 80652 CGGCGCGG 80711

Qy 1088 TCACCGGAGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1147
Db 80712 ACATCGGCGGCTGCGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80771

Qy 1148 TCGGCTCCCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1207
Db 80772 CGG 80831

Qy 1208 CCGAGGACTTACCCCGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1267
Db 80832 TCGGCTCCCTTACCCCGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 80891

Qy 1268 GGCATGATGCT 1327
Db 80892 GGCATGATGCT 80951

Qy 1328 AGTACGCGG 1387
Db 80952 TCGG 81011

Qy 1388 TCGTCCGCGG 1445
Db 81012 ACGCGCTCGG 81071

Qy 1446 CCGCGTGGCGG 1505
Db 81072 CCGCGTGGCGG 81129

Qy 1506 CGAGCGG 1565
Db 81130 CGAGCGG 81189

Qy 1566 CTGGGACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1622
Db 81190 CCGCGG 81249

Qy 1623 GCACACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1682
Db 81250 CGAGCGG 81309

Qy 1683 CGCGCGGCTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1736
Db 81310 CGCGCGGCTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 81369

Qy 1737 GCGCGG 1796
Db 81370 GCGCGG 81429

Qy 1797 GCGCGG 1856
Db 81430 GCGCGG 81489

Qy 1857 GCTCTACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1904
Db 81490 CACCGCGG 81537

RESULT 9

AX211739

LOCUS

AX211739

DEFINITION

AX211739

ACCESSION

VERSION

AX211739.1

KEYWORDS

SOURCE

ORGANISM

Streptomyces noursei

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Streptomycineae; Streptomyces

REFERENCE

1 (bases 1 to 125401)

AX211739 125401 bp DNA linear PAT 06-SEP-2001
Sequence 35 from Patent WO0159126.

AX211739.1 GI:15523950

Streptomyces noursei

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomyces

REFERENCE

1 (bases 1 to 125401)

AUTHORS Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
Gulliksen,O.M.
TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 35 16-AUG-2001;
Norges Teknisk Naturvitenskaplige Universitet (NO); STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO);
ALPHARMA AS (NO); Sinvent AS (NO); Zotchev, Sergey Borisovich
(NO); Sekurova, Olga Nikolayivna (NO); Fjaervik, Espen (NO);
Brautaset, Trygve (NO); Strom, Arne Reidar (NO); Valla, Svein
(NO)
FEATURES Location/Qualifiers
source 1..125401
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"
BASE COUNT 15664 a 49692 c 42871 g 17174 t
ORIGIN
Query Match 7.4%; Score 147.6; DB 6; Length 125401;
Best Local Similarity 45.4%; Pred. No. 4.6e-08;
Matches 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;
QY 8 ACCCGACTTCCCTCGAGGACACCGTCCCTTACGGCCGACCGCGCGCGACTGCG 67
DB 80064 ACGGGCGGACACCGACCGCGCGCTTCGGAGCTACACCGCGCTGTGGACGCGCAC 80123
QY 68 TCCTCAGCTAGACGGGCGCTTCGAGGTCTTCGTCGGGACGCTGAGGCGACTGGT 127
DB 80124 AACAGCGCGCGCTTACGCGGACCTTCGGCGCCATCAGCGCGCGCGCTTCGGCT 80183
QY 128 CGAAGCTTGTCTCTTGGGGCGCGGTGTGACCGGAGCTTCGAGACGGACGCGCGAG 187
DB 80184 GGAAGGCGTCTCGCTCGCGCGCGCGGCGCACACCGCTCGCGCGCGGATCGCGCG 80243
QY 188 GAACGCTGACGTCACCGGATGACGTCGAGGCGCGGACGATGCTGACGCGCTTGAC 247
DB 80244 CCGGCGGAGGACACCTCACTACGCGCTCTACGAGCGCGCGCGCGCGCTGCTGTCG 80303
QY 248 CTGCGCGCTTCTGGGCTGACAGCGCTCCACCGGACGCTCTTACCGCTGCC 307
DB 80304 TCGACTTCTTGTCTCCGCGAGTTCCTCGGAGCACCGCGCGCGCGCGCGCTTCTTCC 80363
QY 308 CGAAGCGGTGAGCGTCTGTCGCGCGCACCGCTCCAGAGAGAACGCAATGACGC 367
DB 80364 ACCGCGACTCCCTCTTCCAGCTGAGTGAGCCCGCTCCAGGCGCGCGCGCGCGCAC 80423
QY 368 CGGCGATCTACGCGCGCGCTGTCAGGAGCGCGCGCGCGCGCGCTGCTGCTGATCT 427
DB 80424 CGGCGA-----CCGTCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80474
QY 428 TCGCGCGTGTCTGCGCGCTGCGAGCGCGGATCGCTCGCGGCTGCTGCTGCGCTGCT 487
DB 80475 TCGCG-----CCAGCGCATCGGACACCGCGCGCGCGCGCGCGCGCGCGCGCTG 80525
QY 488 ACCCATGACGCGCGCGCGCTGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 547
DB 80526 CCGAGCGGAGGCGCGCTGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80582
QY 548 GCGCTTACAGTGTACTACCTGCTCCGACCAAGAACGCGCGCGCGCGCGCGCGCGCG 607
DB 80583 GCGCGCGCTTCCCG 80642
QY 608 ACGGAGGACGACCGCGCGCTTCAGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 667
DB 80643 AGTGCTCGCGGACGCGCTTCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80702
QY 668 CCGACTTCCCGTGTGTCGCGGTGCGGCGTGTGCGGACCGCGCGCGCGCGCGCGCGCG 727
DB 80703 CCGAGGCGACGACCG 80762
QY 728 GCGCGCGCGGTGCTGTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 787

DB 80763 AGAACCCCGCGCGTTCGCGCTTCTCGACTCGCGCGCGCGCGCGCGCGCGCGCG 80822
QY 788 AGACGTACTCTTACTTGGTCGACCGACGCGGGTTCACGTTTACCGCTTCCCGCGCGCG 847
DB 80823 AGACCTTGGCCACCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80879
QY 848 TCATCTCAACACCGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGT 907
DB 80880 GCACGAGCTGACCG 80939
QY 908 GGTTCGCGACCCCAAGATCCACTGGGACACCGCGCGCGCGCGCGCGCGCGCGCGCG 967
DB 80940 GGAACCCG 80999
QY 968 GACGACTCGCGTACCGCGGTTCCTACACCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1027
DB 81000 CCGCGCGCGTGTGCG 81059
QY 1028 GCACTTTCGACTACCG 1087
DB 81060 CG 81119
QY 1088 TCACGCGGACG 1147
DB 81120 ACATCGCGCGCTGCG 81179
QY 1148 TCGGCGCTCCCGATGACGTACGCTTCTGACAGCGCGCGCGCGCGCGCGCGCGCGCG 1207
DB 81180 CG 81239
QY 1208 CGGAGCGCTTACCG 1267
DB 81240 TCGGCTTCTTACCG 81299
QY 1268 GCGCATGATCGACG 1327
DB 81300 GGCACCTGCACGAGCG 81359
QY 1328 AGTACG 1387
DB 81360 TCG 81419
QY 1388 TCGTCCGCGAGCTGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1445
DB 81420 ACGCGCTTCCCG 81479
QY 1446 CCGCGTGGCGCGCTGACGAGTACGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1505
DB 81480 CGTGCG 81537
QY 1506 CGACG 1565
DB 81538 CGCGCGCTCG 81597
QY 1566 TGGGACG 1622
DB 81598 CCGTGGCG 81657
QY 1623 GCACACGAACTCG 1682
DB 81658 CCAGGCG 81717
QY 1683 CGCGCGGTACTGCG 1736
DB 81718 CGCGCGGTCTCG 81777
QY 1737 CGCGCGGTACTGCG 1796
DB 81778 CGCGCGCGGGAAGCG 81837
QY 1797 CG 1856

Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.

1 (bases 1 to 12744)
 Schauwecker, F., Pfennig, F., Grammel, N. and Keller, U.
 Construction and in vitro analysis of a new bi-modular polypeptide
 synthetase for synthesis of N-methylated acyl peptides
 Chem. Biol. 7 (4), 287-297 (2000)
 20243879
 MEDLINE
 PUBMED
 10780924
 REFERENCE
 2 (bases 1 to 12744)
 Schauwecker, F., Pfennig, F., Grammel, N. and Keller, U.
 Direct Submission
 Submitted (12-NOV-1999) Institut fuer Biochemie, Technical
 University Berlin, Franklinstr. 29, Berlin 10587, Germany
 Location/Qualifiers
 1. .12744
 /organism="Streptomycetes chrysomallus"
 /strain="ATCC 11523"
 /db_xref="ATCC:11523"
 /db_xref="taxon:1899"
 1. .12744
 /gene="acmC"
 1. .12744
 /gene="acmC"
 /note="ACMS III; peptide synthetase"
 /codon_start=1
 /transl_table=11
 /product="actinomycin synthetase III"
 /protein_id="AAP42473.1"
 /db_xref="GI:728478"
 /translation="MTSGLEDILPLSPMQEGLLFHSRYEQDQADYAVYVQHVLDLQAP
 LDGARLEAVGALVQRHNPRAAFQVDSQRTVOLISRHVDLPEWFEFSSMPAPAE
 AELSRIAKEHGRFPDPAEPFLRFTLVRTAEHHRLLITLTHHLDGNSPTILVREL
 TALYADGPAAELPRVTPYROYLGLAQODPRAEAAWREALGLSEOTLVAPVDPARA
 ALMPERTTELAERTAAALVEMARRHCNTLVLOAAWGLVLSRRTGHDPVFGAVAA
 GRPQLAGVESHVGLLITWVPYVRDLHDAGSLLEAVRRQLDEQSLTSHHGIARIQ
 YLRTGMSLEFDSLVFENYPMQDPAELPATGLRITPDLGRGRDATHYPLTLIAAPGRRL
 YLRLDGRGLDRAEACALVERLIRVLDLAVADAERPLGRVOLLTAERELLHAGNG
 AGMPLAASPLVFLRAVAAADAVAVEGEALRYRELDERANRDLAHLIAHGVGPE
 RIVALAPSRVLDVAVLAKAGAAVLPDPEYPANRLAHMVTDOHTLALHTTETE
 AKLPDRHGLLDDPAVLADSLGRPAHPVVELHPDHPAVVITSGSTGPKGVMPA
 GGLNLLOWHRRVABEGTRTAQTAISFDSVAQEVLFVAFGKTLVITPDEVRDA
 ARFAGMLDQDELPAFNLALAEAAVETGRTLPQLRTVAQAGEALVLSLTVRAF
 HRSAPGRLLNHYGPTTETHVYTAHALGDPPDWRPLPAPIGRIDNTHAYVTRAVRL
 EPGVGLGYTAGLAGLARGPLPAERFADVPYGLPGGRMYRTGLDVRNPDGEL
 EFCGRADHGVKGRFIEPCEIEKVLTDHPDIAQAAVTRPHRPGDRLVAVVVGREA
 LRPEQVTRERLPRHMPAAVQVLERLPTNGKLDRAALPEPDEALAGTGREAT
 PQQIVCDLFAQVGLPWGVDDDFELGSHLSLRLTLRIARAAVEGLRLTLEA
 RTAAVAHLLDLAGPARTALTQHLPDAVPLPSFAQRRLMFLHKMGPSATYNTPLAIR
 LTGELNLEALRALEDVTRHESLRTVFPEVDGTPYQVRLDTATIELVTRATTAEALT
 DHLRAARHPFDLAGPPLRTFELSAREHVLVVHHIAGDWSLGLPLASDLARAY
 LPTRDRPAPVMTYRGDYLTVDIDPELHRLTELARGSGASLFWVLQAGLAALIKRLGA
 GDDIPGLSPTRDQALDILIGFVNTLVLRDITDGTGFTTQLLTTRTSLAAHAT
 QDVFFELVEMNPTTRTAAHLHFLQIMLQNAPEFTFPLGLTVDVAPQRTAKPD
 IDLSLAERGADEGPGITGAVEYSSDIYDAPVQALFNWRIHLLDTATAQPEQLSR
 IDLLTGAHQHLLDWTDAVEYDPDLPARFARQAATPOAVALLAGDSDSLYAEILD
 ANRNLHAALLREGAGPDRVALAPLTAELVALLVALLKTAALYLPDPDHPAARLS
 HVLGDAPALLLTDTEOHLPADATRLRLASAEVRLALADCPDTPDAEYGTAP
 GSAAYVYTSGSTGRPKGVVPHSALVNFVIMRQAPLRPQERLLATVTTIDIAL
 ELVHPLLSGAAVLAKPAVQPSAVLDLRIAGHTVTVOGTPSLMOLLVGHDAEARLG
 LRMVLGGLTSLAELRALTDLVNLYGTGTTTWTAAELAGTGAAPRIAPRLN
 TRYVVLGDLPQVAVGVGELYIAGLAGLARYLDRPALTAERFPADYGLPEGGRMYR
 TGDVLRWNPDPGVEFFGRADHGVKVRGERIEPCEIEKVLTDHPDIAQAAVVEDPDG
 DARLAVYVTGSADARDEQVQDLDEWQDLVDSYTTASETAFAENFASWNSVDG
 LPILQDMERDWRDITVERIGNPRVLEIGVGTGLLSRLAPHCSEYWGTDSPVIT
 ADLURHVEADPELAAVQLRQTPAHDFDQLPHGHFDTVLNSVYQFPNAGYLEQVID
 HALRILAPGTVIGDIRNPRLLRTTAVHTARATDPADHAAVRHAEHALVLEKEL
 LIDPEYFTALGHVDPDLAVGVDIRLKPCTVENEUTRYDATLHKAGANTHPLGPRER
 PWSGHADPOAELAAALQOLDSERPARLVGTIPNRLARELALQHALDTATAPASP
 TADAGQDPDLALHRLGRHOGYWTATNTHRPDAVLDLVALSDDGATPTGTAS
 AGTSTPATPLUSSLTTPAAGRTSALLTLRREHARHLPDYMQPSALPLDLPLPTAN
 GKDLRALPDLPTAGTREPRTQEQIVCDLFTQVLGPRVIGIDDFELVGHSHLL
 ATRLIAQIRALFGLVELESLFEGPTPAVAARLDTAGPCRLALTVOERPATWPLSLFA

QRLWFIHKMEGSPATYNIPLALRLSGDLDRDALRGALTDLVSRHRESLRTVFPVDT
 PYQLLTPAARLTVPTTSEADLPDAEAAARYADLAEPPLPTELTEFLSAREHV
 LLVHHIAGDWSLGLASDLARAYAAEAGRAPAWPLVQYADYTLWONELHGDQ
 NPDLSLPATQVAYMTETLAGLPQITLPTDRPRPVMYRGDYLTVDIDPELHRLTE
 LARGSGASLFWVLQAGLAALLKHAGDDIPGLSPIAGRTDQALDILGFYFVTVLR
 TDGTGPTTQLLRVRELSLAAYHLEVLNTRTLAHPHLPQIMLALON
 LAGTSPALVDPALVSCATLSYSELNTRANHLAHLTTCRIGPCDAVAVLQSPQTV
 EQOVLTPDPAVQRTGAKFDLFLSAPERRGADGEPQITGAVEYSSDIYDAP
 APEGTQLPGLTVDVAPRTGAKFDLFLSAPERRGADGEPQITGAVEYSSDIYDAP
 TQALFNWRIHLLDTATAQPEQLSRIDLVLTPEERNRTIVEVNRTELPLPDASLAE
 TVYLAALAKTATYIPLDSRYPADRYLRLDTEKLLITDHTTLDLTTQFPADPT
 HDGEDPGNHNTHPDDAAYIMTSGSTGRPKGVIAITHRNITALALDPRDPTAHRV
 LLHSPATFADASTYEIWPLNGNTVVLAPTDLDVHTYVHRVITDQTLTALWLSWVFN
 LLTEQSPETFTVRVQITGEEAVSGATVTRLOACPDVTVDGTEPTTTFATHHPV
 ERFVANYAAGPMYRTGDLARNVDHLEYAGRADHGVKVRGERIEPCEIEVNLTD
 PTFYVTPGSAVPIGRPMATHTYVLDSDLPVAPGVGTGELYLAGSLRGVLDRLPALTA
 EFGEDSGNNSYDGRPIPLDOMREWRDATVERIGLNPRVLEIGVGTGLLAKLAP
 HPAVAAVHLNRDPCNPRLVAVVADTSAPSSDDVQOHOIGEMODLXSLVAAPTA
 VOYFNADYLAQVIEQALRLPAGGAVFIGDIRNPLRLTFTTAVOTARADPADTAA
 VRRVAVQSLVLEKELLDPPEYFTALHRLPDLAGVDLRLKGAHNLRTYRYDITLH
 KAGITAPLSEAAVLAWPQDAELARHLAEARPERLVRTGAPNSRLAADIQAHALES
 GTAPAGPTGYPATEQPDLEALHRLGEDHGYWTAVTWSAHRPDTVDLTFVRGLLDGA
 VPYGTYAPAAAGDPATPLAFTTNPVSGRTAALLTALREHAAQLPDYMRPAIVPL
 DRPLTANGKLDRAALPALDPEHADTGRAPTPQEVVCELFVLEGRPLVGVQDQFF
 DLGHSLLATRLIARLRAAFVGLRSUFEPAPTPGGIAARLDLDDPGSYEVVLPUR
 AQSRRPPLFCIHPPGGISWSYSAIKHLGPEYPLYGIAQARSARLPFRPESIEEMAVD
 YADQIOGVQPHGPGYHLAGSGGICAHAAAEFORGEVALVAVLDVIPWQGLTTHD
 DVPADPDVRLVYHGLVDGSHRNDREELTFAAREILRROGSLANLEEDRLTIT
 EISANNTHLTVQPGPIDGDLILLIASEQDDPPVTADAWRPYVCGAVEAHVYVGEH
 SMUTRHTLAEIRIGLISAKULELDGA"

BASE COUNT 1846 a 5446 c 3884 g 1568 t
 ORIGIN

Query Match 6.08; Score 119.8; DB 1; Length 12744;
 Best Local Similarity 44.68; Pred. No. 9.6e-05;
 Matches 558; Conservative 0; Mismatches 687; Indels 6; Gaps 2;

QY 583 AACAAACGCGCGCGGTGGAGCACGACGACGACGACGCGGTGCGCTTCCACGCAC 642
 DB 9049 AGCGACGCGCGCACGCTACTCGAGCTCAACGCGCGCACCACTCGCCAC 9108
 QY 643 CACGCGACCGGTGATCGCGCTCGCGCGCGGTGCTGCTGCGCTCGCGCGGTGCTG 702
 DB 9109 CAGCTCACCGCGCGGTGATCGCGCGCGGTGCTGCTGCGCTCGCGCGGTGCTG 9168
 QY 703 GGCACCGGACACGCGCGGTGCTGCGCGCGGTGCTGCTGCGCTCGCGCGGTGCTG 762
 DB 9169 CCCGACCGGTGACCGCGGTGCTGCGCGCGGTGCTGCTGCGCTCGCGCGGTGCTG 9228
 QY 763 ACCGACCGGTGCGCAAGTACGAGGACGAGTACCTCTACTGTCGACCGCGCGGTGCT 822
 DB 9229 GACAGCGGTGACCGCGGTGCTGCGCGGTGCTGCTGCGAGGACCGCGCGGTGCT 9288
 QY 823 ACCTTACCGCGCGGTGCTGCGCGGTGCTGCTGCGAGCGGTGCTGCGCGGTGCTG 879
 DB 9289 CTCATACCGCGCGGTGCTGCGCGGTGCTGCTGCGAGCGGTGCTGCGCGGTGCTG 9348
 QY 880 ACGCCCGCGGTGCTGCGCGGTGCTGCTGCGAGCGGTGCTGCTGCGAGCGGTGCTG 939
 DB 9349 ACCCGCGCGGTGCTGCGCGGTGCTGCTGCGAGCGGTGCTGCTGCGAGCGGTGCTG 9408
 QY 940 GCGCGCGGTGCTGCGCGGTGCTGCTGCGAGCGGTGCTGCTGCGAGCGGTGCTGCTG 999
 DB 9409 GCCTACATCATGACAGCGGTGCTGCGCGGTGCTGCTGCGAGCGGTGCTGCTG 9468
 QY 1000 CCGAACCTCGCGGTGCTGCGCGGTGCTGCTGCGAGCGGTGCTGCTGCGAGCGGTGCTG 1059
 DB 9469 CGCAACATCACCGGTGCTGCGCGGTGCTGCTGCGAGCGGTGCTGCTGCGAGCGGTGCTG 9528
 QY 1060 GGCATGAGTGGCGGTGCTGCGCGGTGCTGCTGCGAGCGGTGCTGCTGCGAGCGGTGCTG 1119
 DB 9529 CTCCTCACCTCGCGGTGCTGCGCGGTGCTGCTGCGAGCGGTGCTGCTGCGAGCGGTGCTG 9588

QY 861 CGAGCTCGGGCGGCGCCACACCGCCCGGAGATCGAGAACGCGCGAGTGTTCGGCGACC- 919
Db 1671 CGCGGTCTCAATCGGGCGCGCTTACCTGCGGTCTCGACCTACCGCGCGCGACCG 1730
QY 920 -----CCAAGATCCACTGGACACCGCGCGGAGAAATGGTCTCGTATCGGACGACT 974
Db 1731 GCTGGCTCGATCTCGAGACGCGCGCCCGCTCTCTCGACCGCGCGGATGAC 1790
QY 975 GGGGTACCGCGGTTCTACACCTCGCGAAGCTTCGCGACTGGACACTTCGCGCGAACTT 1034
Db 1791 CGAGGACCTCGCGGCTACCGGACACCGCGCGCTACCGCGCGCGCGCGCGCGCG 1850
QY 1035 CGACTACCGACCGCGCGCTCGCGGATCGAGTCCCGCGCGCGCGCGCGCGCGCGCG 1094
Db 1851 CGCCTACGCTATCTACACCTCGCGGCTCGACCGCGCGCGCGCGCGCGCGCGCGCG 1910
QY 1095 AGAGGACGACCGCGCGCTCGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1154
Db 1911 CAGCAAGCTGGCGGCTTTCACCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1970
QY 1155 CCCCATGAC---GTACGCGCTACTGGACGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1211
Db 1971 CGTGTGACCGCTTCTCACTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2030
QY 1212 CGACCTACCG 1259
Db 2031 CCGTGTGCG 2090
QY 1260 CGTCACTCG 1319
Db 2091 CTTCTCGACCTGTGGCG 2150
QY 1320 CAACGTGAAGTACCG 1379
Db 2151 CCACGAGCTCG 2210
QY 1380 GAACGTGATCG 1439
Db 2211 CCGTGTGTGGCG 2270
QY 1440 GAGCACCG 1499
Db 2271 CGCGACG 2330
QY 1500 GACCGTTCGAGCG 1559
Db 2331 GCACG 2390
QY 1560 CATCGCGTGGGACCG 1616
Db 2391 CATCG 2450
QY 1617 AACCGCGCACG 1676
Db 2451 CACCG 2510
QY 1677 CGACCTCG 1736
Db 2511 GCTGACCG 2570
QY 1737 CGCGCGATCGGCG 1796
Db 2571 CACCG 2630
QY 1797 CGCGCGCGCACG 1856
Db 2631 CGACCGGTCAGATCG 2690
QY 1857 GCTCTACCG 1916
Db 2691 CACCG 2750
QY 1917 GCGCAT 1922

Db 2751 GCGCAT 2756

RESULT 14

SPSNBCGEN

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

gene

CDS

SPSNBCGEN 22449 bp DNA linear BCT 24-MAR-1997
S.pristinaespiralis snbc and snbDE genes.
X98690
GI:1906377
Pristinamycin I synthase 2; Pristinamycin I synthase 3 and 4; snbc
gene; snbDE gene.
Streptomyces pristinaespiralis.
Streptomyces pristinaespiralis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 22449)
de Crecy-Lagard, V., Blanc, V., Gil, P., Naudin, L., Lorenzon, S.,
Famechon, A., Bamas-Jacques, N., Crouzet, J. and Thibaut, D.
Pristinamycin I biosynthesis in Streptomyces pristinaespiralis:
molecular characterization of the first two structural peptide
synthetase genes
J. Bacteriol. 179 (3), 705-713 (1997)

J. Bacteriol. 179 (3), 705-713 (1997)
97158664
9066024
2 (bases 1 to 22449)
de Crecy-Lagard, V.
Direct Submission
Submitted (21-JUN-1996) V. de Crecy-Lagard, Institut Pasteur, 25
Rue de Dr. Roux, Paris Cedex 75724, FRANCE
Revised by [3]
3 (bases 1 to 22449)
de Crecy-Lagard, V.
Direct Submission
Submitted (24-MAR-1997) V. de Crecy-Lagard, Institut Pasteur, 25
Rue de Dr. Roux, Paris Cedex 75724, FRANCE
On Mar 25, 1997 this sequence version replaced gi:1483602.

Location/Qualifiers
1. 22449
/organism="Streptomyces pristinaespiralis"
/strain="SP92"
/db_xref="taxon:38300"
/clone="pIBV3 cosmid"
115..7890
/gene="snbc"
115..7890
/gene="snbc"
/codon_start=1
/transl_table=11
/product="Pristinamycin I synthase 2"
/protein_id="CAA67248.1"
/db_xref="GI:1483603"
/translation="MLVPDDARLPDLTAQLGWFQAQRDPHSPVNTGEVVELTGGGLD
ADREALRATVARTDALRFRHTDAEGTQTQSVADDDVFFTLHRLVDVSAEDDDPAAAD
AMRDVIRLDVTCGRPLETQALFTGPGHRWFQHVHLLLDLGGYGRFLVAERAAA
YRAGAEGDAEAGDETPGIRDLAAEDWASDAYAQDRAHWTAPRHTPVTLAGTTA
PGAPCALRRTASAGPEVLGLDVAARAAETQROVAFPAVAAYVOMLTGERTVTLGMP
VTARTSAARRAPCTVANVPAHVPEPPAPLAELVASVRDALEPLLRHORYPYEAR
RELGNASERAYGPTVNVNPGYRALDFCGPAAHVHPSTGVPEDLKFVYEBPDGSGAR
IDLEANAAYGESETDAHLRFLVFLGAFAGADPTPLGRTHAAAPAPAAAPAAAPTGG
PALHTLFAARAAATPHAVVSHDGHLLTYAQLDANRANRHLRVEQAGPRLVALTL
PRSLDLVVALAAVLKSAAYLPLDPDYADRLAWMLDEAPVLVLDPOAMTEDLAGYP
DTPRTAVDGAHPAYVTYSGTSGPKGVVTPSHSNVRLFTSTOHWFGEFGDDVMTLF
HSYAFDSVWEIAGLLGGRLVVVPHVSPGDFLDLLAREKVTYLNQPTAFHOL
DAADRATAPELALRYVFGGEALDVADYARBGTAARLVNMYGTTTTHVTH
APLPGHNAVPGTSPSLGGIPDLTPVLDLADALRPVPGFTGELVAGLAGLGNRNP
ALTQGRFPADPYGAPGTMRYTGDVLRHLDGTAYLGRGDDQVKGIRFRELGEIEN
VLATHPGVAAQAAVYREDRHDGLRLAAYVPTGTEPDVAGWRALLAGLPAHSLPAS
FTLLAALPLTANGKLDRAALPAQAPAPAPADEPAPADGCEALRLRFADTLGTPRV
GADSDFTTGGHSLLAGRLIAAVRDTLGAETIQDLFEAPTAGLAARVRAATYATGPRP
LVPRDHGDPPTSYAORHMYVREVEGAAPTYNMLALRLSGTLNRTLAALYDVV
VRHESLRTVPRGHEGVTPLPPPEQARPWIESRVDEQPLAALTAARPEFNLDCQ

QY 333 GCCACCCGCTCCACGAGGAGAACACGATGACGCGCGGCATCTCACGCGCGCGCTG-- 390
Db 1131 GGTGGCGAAGCGGCTACGGCCCCGTGTCACAGCTATGCGGCGCGCGCTGGA 1190
QY 391 ---CTCCAGGAGCGCGCGCGGAGCACTCGCCCTGATCTTCGGGGGTGCTGCGCGCT 447
Db 1191 CTTTGGGCGCGCGCGGCTCGTGCACCCGCTGTCACCGGCGCGGTGGAGACTGNA 1250
QY 448 GCAGCCCGGGATYCGCTCCCGGGCTGCTGCTGCGCTTACCACATGACGCCCGCCAGC 507
Db 1251 GTTCTCGTGTACGAGGAGCGCGGCTCGCGCGCGGGATCGACCTCGAGGCGCAAGC 1310
QY 508 GCGTGGCT-----CTGGACCCCAAGCGCGCGGTACACCCAGCGGCGCTACAGCT 560
Db 1311 GCGCGGTACGCGAGGAGAGACGCGCCATCTGCGCGCTTCTGCTTCTCGG 1370
QY 561 GTACTACTGTGACTCTCGACCAAGAACGCGCGCGGCTGGGACACGCGAGACGAC 620
Db 1371 GCGCTTGGCGAGGCGCGCGGACACCGCGCTCGCGCGATCCAGCGGCGCGCGCGC 1430
QY 621 CGAGCGCGTGGCTTTCAGCACACGCGACCGTGTGCGCTGCGCGCGCGCTTCCCGGT 680
Db 1431 CG 1490
QY 681 GTGCTCGCGGTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 740
Db 1491 CG 1550
QY 741 CGTGGCTGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 800
Db 1551 CGCGCGCTGAGCG 1610
QY 801 CTGGTGCAGCGCGCGGTTCACGTTTACCGCGCGCGCGCGCGCGCGCGCGCGCG 860
Db 1611 CCGCGAAGCGCTGCTGCGCGCTCACCGTTCGCGCGCGCGCGCGCGCGCGCGCG 1670
QY 861 CGAGCGTGGCG 919
Db 1671 CGCGTCTCTAAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1730
QY 920 -----CCAGATCCACTGGGACACCGCGCGCGGAGATGGGTCTGCTCATCGGACGACT 974
Db 1731 GCTGGCTGGATGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1790
QY 975 GCGGTACGCGCGGTTCACGCTCGCGGAACTGCGGAGCTGGACACTTCGCGCGCAACTT 1034
Db 1791 CGAGGACCTCGCGGCTACCGCGACACCGCGCGCGCGCGCGCGCGCGCGCGCG 1850
QY 1035 CGACTACCGGACGCGCGCTCGCGGCGATCGAGTGGCGCGCGCGCGCGCGCGCG 1094
Db 1851 CGCCTACGTCATCTACGCTCGCGGTCCACCGCGCGCGCGCGCGCGCGCGCGCG 1910
QY 1095 AGACGACGCGGACCGCGCTGCGGTCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1154
Db 1911 CAGCAAGCTGGTGGCGCTTTCACCTTCCACCGACACCTGTTGCGCTTCGGCGCGCG 1970
QY 1155 CCCCCATGAC---GTACGCTTACTGAGCAGGCACTTGGAGCGCGCGCGCGCGCGCG 1211
Db 1971 CGTGTGGACCGCTTCTTCCACTCTACGCGCTTCGACTTCTCCGCTCGGAGATCTGGGGCG 2030
QY 1212 CGACCTACCGCGCAATGGCTCG-----ACTGGGCTGGGACTGGTACGCGCG 1259
Db 2031 CTTGCTGACG 2090
QY 1260 CTTCACTGGGCGATGATGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1319
Db 2091 CTTCTCGACCTGTGCGCGCGGAGAGAGTCACTCTCTCAACAGACACCGCGCGCTT 2150
QY 1320 CAATCTGAAGTACG 1379
Db 2151 CCACAGCTGACG 2210
QY 1380 GAACCTGATGCTCGCGGAGCTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1439

Db 2211 COTCTGTTGGCGCGGAGGCTCTCGAGTGGCGCGCTCTGCGCGACTGGTACGCCGCGG 2270
QY 1440 GAGCACCCCGCTGGGAGGCTGACGAACCTAGCTACCGCGCACACACACTCCCGGACG 1499
Db 2271 CGGACAGCGCGCGCGCGCTGGTCAACATGTACGGCATCACCGGAGACACCGCTGACGTAC 2330
QY 1500 GACCTGACAGCGCGCGCGCTGCTGCGCATGGAACGAGCGCATAGGAGATGACGCTGA 1559
Db 2331 GCACCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2390
QY 1560 CATCGCTGGGACACGCGGACAGCTGCGCATCTCGGTGGCGCG--TCCCGCGAGCG 1616
Db 2391 CATCGCGGACTCACCGCGCGCGCTTGGACCGCGCGCTTGGCGCGCGCGCGCGCTT 2450
QY 1617 AACCGCGGACAGCAATCGGCAAGTAGGAGCAGACCTGTACGTGACGAGGACCTCTC 1676
Db 2451 CACCGCGGAGTGTAGCTGCGCGCGCGCGCGCTGCGCGCGCGCTACTTGAACCGCGCG 2510
QY 1677 CGACCTCGCGGCTACTCGCTGCGCGCGCTTACTGCGGAGCGCGCGCGCGCGCGCG 1736
Db 2511 GCTGACCGCGGAGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2570
QY 1737 CGCCCGATCGTGCACCTGCGCATCTCGTGCAGACCGCGCGCGCGCGCGCGCGCG 1796
Db 2571 CACCGCGGACCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2630
QY 1797 CGCGCGCGCACCGCTGCTTCCAGAGCTTCCAGTTCGCGGAGGCGCGCGCGCGCG 1856
Db 2631 CGACGAGTCAAGTTCGCGGCTTCCGATCGAATCGCGGAGATCGAGAACGCTCTCG 2690
QY 1857 GCTTACACCGAGCG 1916
Db 2691 CACCG 2750
QY 1917 GCGGAT 1922
Db 2751 GCGGCT 2756

RESULT 15

LOCUS AF210249 77457 bp DNA linear BCT 29-AUG-2000
DEFINITION Streptomyces verticillius bleomycin biosynthetic gene cluster,
complete cds.
ACCESSION AF210249 AF149091
VERSION AF210249.1 GI:9937210
KEYWORDS Streptomyces verticillius.
SOURCE Streptomyces verticillius.
ORGANISM Streptomyces verticillius
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomyces; Streptomyces.
REFERENCE 1 (bases 74421 to 74693)
AUTHORS Du, L. and Shen, B.
TITLE Identification and characterization of a type II peptidyl carrier
protein from the bleomycin producer Streptomyces verticillius ATCC
15003
JOURNAL Chem. Biol. 6 (8), 507-517 (1999)
MEDLINE 99352421
PUBMED 10421758
REFERENCE 2 (bases 1 to 77457)
AUTHORS Du, L., Chen, M., Sanchez, C. and Shen, B.
TITLE An oxidation domain in the BlmII non-ribosomal peptide synthetase
probably catalyzing thiazole formation in the biosynthesis of the
anti-tumor drug bleomycin in Streptomyces verticillius ATCC15003
JOURNAL FEMS Microbiol. Lett. 189 (2), 171-175 (2000)
MEDLINE 20389599
PUBMED 10930733
REFERENCE 3 (bases 1 to 77457)
AUTHORS Du, L., Sanchez, C., Chen, M., Edwards, D.J. and Shen, B.
TITLE The biosynthetic gene cluster for the antitumor drug bleomycin from
Streptomyces verticillius ATCC15003 supporting functional
interactions between nonribosomal peptide synthetases and a

```
polyketide synthase
Chem. Biol. 7 (8), 623-642 (2000)
20500448
MEDLINE
PUBMED
11048953
REFERENCE
4 (bases 1 to 77457)
Du, L., Sanchez, C., Chen, M., Edwards, D.J. and Shen, B.
Direct Submission
Submitted (01-DEC-1999) Chemistry Department, University of
California at Davis, One Shields Avenue, Davis, CA 95616, USA
On Aug 29, 2000 this sequence version replaced gi:5326869.
COMMENT
Location/Qualifiers
source
1..77457
/organism="Streptomyces verticillus"
/strain="ATCC15003"
/db_xref="ATCC:15003"
/db_xref="taxon:29309"
complement(1199..2185)
/codon_start=1
/transl_table=11
/product="syrp-like protein"
/protein_id="AAG02342.1"
/db_xref="GI:9937211"
/translacion="MTGSLTLPGLGIIIPRGEGLTTGAEXDGLGLDAGPDWVRAH
GPRLERLATDGLLHGLPTDGDGVDGFDHVGSGVDGGLPPTERSTPRSVKGNLY
TSEYFADQIPHNENSYAAHWPSTLYFFCHTAPDGGAIPIADGRAVLDLIPAEVR
RFSQGVYTRTFADMGLSWQAFQEDRGDVERHCRHQEQFSWDDGDLVLRTHHRP
ATAVDPTGAEVWFNOAHLPHPSLSDPLRQLVLETYGNGLPDGLFADGCTIPDAD
LATVRAATRAALAPWREGDIMLVNDRMAHGREPTGERRVILVAMTSADS"
complement(2182..3918)
/codon_start=1
/transl_table=11
/product="peptide synthetase NRPS12"
/protein_id="AAG02343.1"
/db_xref="GI:9937212"
/translacion="MTLTLRDAFLDQAARTPDAAHVVGDTVMTYRELELRAGRMART
LAARGPGTLVAVLRPGPEPVAALLAVILTGAGYPLADDDPPDCRRLHLDCAA
LLAEHPROGRITLDEALAPARPDAAVPRAGDPAYVIITSGSSGRHILVEQGA
LGAYLAQARYDGLSGRTVHLSLSFDMATSLMGLVSGGAIHVLDLKAIASGTOP
PAAASRPSFLKVTYPSHLPLGLIPDSCLTQGLVIGGELTSGALGPWRAAHPDVTY
VNYGPTAEATGCAVTRPGDAVDPGAIPIGRFAGTLYVLDAADGEPVAVGGVGL
HAGDOLARYLGRPLRTERFVDPDPAAGSRMYRTGDLVREPDGDLVLEGRADGQ
VKVSYEIEGETEVLHGHVGRDCAVAVGEADARLVAIVYVDPDPSPGCTAAPAR
HAAELPPYWPATFVTYVPELTPNGKLDRDALPGPPAGDAGPGDKTPAETILCELL
ARALGPEIDADADEFITSGSTITALKLVAGARRVGIRLETLTVLRERTVRRILAAOP
DAASPLAEGYPE"
complement(3915..4733)
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG02344.1"
/db_xref="GI:9937213"
/translacion="MSKNAHWSRIKRTGDAGVGVVLAVDFTYGTQREATFRHCLDLD
PVEWHAVPAPDGDWSTAGHRLRWTEGLDVTYLAGRPVRLVAGCVGFASALA
DALVREGRHPRVVLNPNFAPGVAATLROFRLGTAGMDLLTDGGERALLAETAIRRA
HAPDALVFAERVAALYREGCDLIERGLDVSFGAELAAVLHSLYLAITLALDVPPT
PLMRGAVSLRSREHQGTDTDFVEHGFDAVARAELLSSPVVAAITALLREHEAR"
complement(4741..6024)
/codon_start=1
/transl_table=11
/product="putative transporter"
/protein_id="AAG02345.1"
/db_xref="GI:9937214"
/translacion="MTARVATPTYLAVLAVREARAPLLGSLCLARMSFVPLPLALLS
VRDTGSFAGLTSGLSALSTLTFAPARALIDRGSRSLVRLTVPYLILGLAVLIT
LAEAEPTALLVAAVAGVPAPPLGPTMRVLRARILHGRQPLHTAYALDSVTEVY
FTVGPLLGGIILVAAAPLASMITVWLIATAAGTACFVLSAATAAPASGADEDRPHGR
PMALPGMRTLVISFGVGLVGVVQLVYLPIADHAGSPGAGGILLSMSAGSVGLLA
GRIAWRSTPVRPNTVLTGFTLAVPLCLTASVPVAGAFALLVGLCLAPLFTTAYLL
VNDLVASGTAPTPEANTVSTANNNGGFAAGSAAAGVLLDSRGPTVTVTVAFAVAATA
VMTVLRRTLLLGGHPPEAAATPADRTAPAEAE"
complement(6021..6773)
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="AAG02346.1"
/db_xref="GI:9937215"
/translacion="MDPLTRVNPWFSGCDGRPRVRLCALPVAGGTAAVFKDWPAAAL
PGVELLTAHLPGRGDRFTEPPATLEETAERLCEALPPSDLPVTVLHSGMALLGYE
VAARLAARGRAPNLLIAACAPPPHPPDASGVTEAEALATLRAERERWDRALRDEEM
EAVLPALVADITAGDRYRPRPLDPLKVIYIGADDDGTDRWTLTGWRACRACDEV
VVLPGGHYFLETDRAAVLTRVATDLAEAEVGA"
complement(6784..7797)
/codon_start=1
/transl_table=11
/product="clavaminic acid synthase-like protein"
/protein_id="AAG02347.1"
/db_xref="GI:9937216"
/translacion="MTDLPLRTVALTGESAEVDDLLRTLADVPDSTVGLLHRTLIA
AOELPLIRAEILTCMRLYDSPRALVTVTGFVDDERIGTPTFAARPAPDPTERDLELL
LLHUALLGEAFGNATQNGRLVHDVLPVPEETAAQMSSETELLWHTEDAFHPLRCD
YVGLLCLRNHQRATTVWPDLSKLITTEDRAVLLLEPRYLIRPOTSHPTPAONATGTRSA
ERFAATAEMDDAPRAVLPEDPEDPYLRIDPAYMSPAGDAARADAYDITVALLIDE
LRHVVLDAAGSLLLVNDNYQAVHGRKPFAAAYDGRDRLKRVNITRDLRRSRARSATS
LLV"
complement(7859..9214)
/codon_start=1
/transl_table=11
/product="putative transporter"
/protein_id="AAG02348.1"
/db_xref="GI:9937217"
/translacion="MKKSRRLRQLGDEPPVVAOESPQDAGTPCQGDGLNVAAL
AALLEVEVPVRPLPHAGLGRHVEHSQAAVALGPMQVASATITGVAGDGHQGRRQ
RGVSHDGLGDVLPQGRGLSLGSLRRRGDVGTAHAGVQPGQPGPAAAECSRG
LVARQDQDQALLRAVVEGPFQGEDAQGVAETVHDHPDVGHEVTVGGQOREVGGOL
GHWDRREVSVAGDGFQGVGLRLAPVSAGHAVDRAAGHVDGLLAVGREGQQQ
QRGGRAGDVHCPLVQGVEDLTHGGODRLIVADLLRPQKTHGLDHRPRVPLARV
H"
complement(9211..17736)
/codon_start=1
/transl_table=11
/product="peptide synthetase NRPS11-10"
/protein_id="AAG02349.1"
/db_xref="GI:9937218"
/translacion="MASDALSAHLESDCQRGGITVNDGSELAGONVAARPERIS
AIAPERTAILHKAATGYDELNRRAELTATLADAGASTPSTLVAVALPRPDVATLIC
ALLKGAACLPDPIGAPRLREIMADAPDLVTTTRAVAPFTGDPGLFLDDAPT
CSAVLPHRSAGTASEIAYLYPTTDEKSENSVSVYRDMARYLDDPTAGTIPARAEILR
LVAPLLSGRLVLDADETRPVPTREAPDMVEDVVAQVCAVLGVDRVGVDRDFDL
GKSLAAVQVVARLKLGLVLPRLALFADPTVEELAAVRAEAGAGQGVREANALPP
VGRSEPLPSFAQORLWFLDRMLPDRAFYTMCDAFVRVGGIDGLALRRALMLVGRHE
TLRTAFYERDGVYQLVPADGFCARRVAAPTRVDLSLLEPAREEAVRNLLVAEAT
PFRPADGALLRVVARLADDDHVLVSTHHIVSDAWSVGLVDELGLRYRECVTGDA
ALPPAVQYDAFVAMWAMWAGPVQEEHAYWKRALDAPSVI RLPMDHPRPAVQSER
GETVGFALPDALVAELEKLGREGOATLMTLLGAFVOLLARHAGOEIDVVGVAAGRT
RTETEPVGFVNTLPRAICAGLSFEDLLDQVREAAALGNFAHQDLPFEALVEALAP
ERDLGNHLPVQVTFQQLGCAFRPDLLIGTEVYPOQEAQVSDLSLDTIRADDGSTR
GIINYGLGTLDRRMEVILVGHYLTLLGAAAADPGRTIELPSGAEARLLDGFGR
DAAYAGPSVPERFEAVARTAPARATCGATLTTFEALNDRVERLAQALGAGVTR
TPAVRLPRSTDSVALLAVRAGGVVPLDPPDPAORTAYILDRTAAVVITRDLPA
LPGLHVDPRPADGLVAPRIDPQAAVITYSGTGAPKVVVRHRSILNHTSAL
QATFLGHDPYLAGADGVPPGDAKLRTLTLTAFDASMEQLSWMLAGHELFTYVEDVR
RDFSALVRFVREHRIIDVITSSQLELLVSHGLDGEWAPSMVAVGEAVSPSLWRTL
RDQRTCFNLYGTEATVDATCDSDPADPVITGLPHTHVRVGLDRLRPVPVGV
AGETYLGGTGLGRVYLNRPALTARFVADPVPTPGSRLYRTGDRARPDGGLTYLG
RTDQIKIRGRFVEPGEIEALVTHHPAVKEAAVDDAHARLVAVTVLAEGGAGPTDV
RRAQGRPLAHMVPVAVVLEALPLTSNGKLDRLAPAPAGRELDVREVPADWVE
EVAAQWCAVLVDRGVGHDFELGSLVLLVQVMTIRKLLGVEVPLRELFDAAVY
BEAARVAAATTEGLGGAAPPLVDSRGLPLSQAQKLWYLDQLAPDSVSNMCD
EYVRGPTDLIDALRLALTRVERHETLRTAFVERDGVPHQVVSAPDAPAAARAEVVR
IEAQRGTDEAROLVAEATPRPADGALMARVARLADDDHVLVTVTHHIVSDGWS
VDLILDELGRLYREHVTGDPAGLPDVOYADFAVMORSMTGTPVREHLYAKNRALD
GAVSLRLPADHPRPAVQSORGETVEFPLPAPVARLEALCREGQVTLFWALLGAFQV
LLARYSGQDDVVGVVPTANTRAETEPLVGVFVNTLPVRVACSPFELSFWALLGAFQV
```

ALGAFADHDLPEALVEALAPERDLGHPLVQVTFQLLDAPDERLVLHGTCVSLGFG
GVTDFSLDSVSRGRKCVLTYPDLDFPRAEVLAGHYLTLGLAADDPLGRVG
DLPSDDVRLRLGSGSPRYLPAPAEVTPDAFVAQVRAFPDAPALVHGSDTLTFAE
LDRTVTAALRLRCGVAETPVAVCLPRSDADVALLAVLRAGGVVVPVDPENPSGR
VAHVLDTEAAVPIVTRDLPADPRVHLDPQAPADDPDLPLRLHQDAAYLIFTSGST
GAPGVVVRHSLHLHGLHVRMAEGGRNVVHTTAMTFDPSLEQFLMVLVAGHTLHV
APEEVRDRPEALVALVRAAIDVLNVTPSHLTLLEAGLEGRVPGTVLVGGRAVPA
ALWRTLRERTGATRFNLYGPTVATDATHLSDPADVPVIGTPLPHTVRVLDDRL
RPVGVAGELYLGCTGLARGYLNRPALTQAOREVADVPDTPGSRVLRGRARWRPD
GTLEVLTGDDQIKIRFERVEPEETEAVLTHHPAVKEAAVTVADDDGAARLVLVPA
PRAPHGSADAPAOVEWNAVFEATHDAADGELFNKGNWDSLTGAPIPAEHMR
EWDVTTVALLREFAERVLIEGSGTGLMLRLPHVTEYTGTDTSRPVADWLDRGR
RPAHRLVLRHEATDFTGVRASATDLVVNSVVOYFPDRAYLDTVLARALDADRGR
VFVGDRLNALQFYARQALAHAGPGAARADVARAAGEFAFAMDEGLIVSPAYFAALA
ARSPIRVTEILPRGRHNMESLYRYDVLHVGGRDPAPEAEVLTWGDVHDLASL
SARLGRGPDALLVRGVANDRLTRDNELLDPARTAVEPEDLMGLADSTPYRVSVSW
AAADPRGMVLLVRRDHDGDLPLVPHVPEPSAPLNTPTPRPSARQSGSADRLG
SWAERLPAHLUPARITEVDALPRTGKDLRGALGLVLTAGRARAGDRPATARTG
LERTLADAWARVLGLPEVGVHENFFALGGDSLLAVRAVARRAGVRLTVROLLSEQT
VAALAAALEESQ*
complement(18307..18660)
/codon_start=1
/transl_table=11

Query Match 5.6%; Score 112; DB 1; Length 77457;

Best Local Similarity 42.4%; Pred. No. 0.00038;

Matches 794; Conservative 0; Mismatches 1065; Indels 13; Gaps 3;

QY 9 CCCCAGCTTCCCTCGACGACACACCGTCCCTACCGCGGACCGCGCGCCGACTGCT 68
DB 38086 CCCCAGGAGCGCGCTGCACAACTGCTCCACCTCGCCCGGGGCTTCGGCGCGGTGA 38145
QY 69 CCTCAGCTAGACGGCGCCCTCTCGAGGTCTTCGCGGGACGGTAGGCGGACTCGCTC 128
DB 38146 GGAGCGCACCCCGCGCGTGTGACCGTGACACCGGTGCCACGAGTCTCGGGA 38205
QY 129 GAACCTGTCTCTCGGGGCGCGGTGTGACCGGAGCTCTGAGACGGACGGCCAGG 188
DB 38206 CGACCTGCCCAACCCGAGCAGCACCGTCCCGCGCGGCAAGTATCCCGCGGA 38265
QY 189 AACCGTGACGCTGACCGGATGACGTGAGGCGCCGACGATGCTGAGCGCCCTGAAC 248
DB 38266 GTACCGGTGGATCGCTTGACCGCTTGACGTGAGCGGGGCTTGACCGCGCGT 38325
QY 249 TCCCGCGCTTCTGGGTGACGAGCGCTCCACCCGACAGCTCTCTCTTACCGTGGCC 308
DB 38326 GCGGACCTGATCGTCCGGAACTCGGCGGCGCGGACACCGTCAACCGCTGCGG 38385
QY 309 GAACAGGCTGACGCTTGTGCGGCCACCGTCCACGAGAGAACAGCAATGACGCC 368
DB 38386 CGGCGGACGCGCTTACACCCCTGCGCGGTGCGGAGCGCCCTCCCGCGGACCGAAG 38445
QY 369 GGCATCTCACCGCGCGCTCTCAGGAGCGCGCGGACGACACTCGCGCTGATCTT 428
DB 38446 CCGCGGCTCGCGCGCGGCGCTACCTCTGTGCGGCGGCTCGGCGGCTCGGCT 38505
QY 429 CGGCGGTGCTGTGCGCGCTGACGCGGCGATCCGCTCGGCGCTCGCTCGCGCTCTA 488
DB 38506 CCACCTCGCGAGTAGTAC---TGGGCGCGCGCGGACACCGCTGCTCTCACCCACCGCG 38562
QY 489 CCACATGAGCCCCCAGCGGTGCTCTGCGACCCCAACCGCGGTCAACACCCAGG 548
DB 38563 GCCCTTTCGCGCGCGGTGAGCGGTGCGCGGACACCGGAGCGCGCT 38622
QY 549 CGCCTACAGCTGTACTACCTCGACTCCGACCAACAGACGCGCGCGCTGGGACCA 608
DB 38623 CTTCCGCGGCTGCGCTCCCTCGCGGACACCGCGCGCTGCTGCTCCGCGGCGCA 38682
QY 609 CGCAGACGACGCGGCTGCGCTTCAAGCAACAGCAGCAGCTGATCGCGCTCGCGC 668
DB 38683 CTTCAACGACGACGCGGATGCGCGGCTTCGCGGACGAGGTGAACAGGCGCGCGC 38742
QY 669 CGACTTCCCGTGTGTCGCGGTGTCGCGGACCGCGGACGCGGAGGTTGCG 728

Db 39813 GCCGCTGCTTCGCCCAAGGAGGCTCTGGCTCCTCCACCAAGTTCCACCCGCGACGAG 39872
Qy 1809 CGTGCTCTCCCAAGCAGGTCCACTTCGCCGAGGGGCGGACACAGGGAATCTCGCTCTACACCGA 1868
Db 39873 CGCCTACAACATCCCCCTGCACATCGCCCTGCGGGGCCCTGAACCGCGCCCTGCG 39932
Qy 1869 CGGCGGCCCCGC 1880
Db 39933 CGCGGCCCTGGC 39944

Search completed: May 10, 2003, 08:41:16
Job time : 6017.5 secs

THIS PAGE BLANK (USPTO)

XX Rhee S, Song K, Kim C, Ryu E, Lee Y;
XX WPI; 2001-308483/32.
DR P-PSDB; AAB82301.
XX Producing difructose dianhydride IV from sucrose, involves reacting
PT sugar solution in the presence of levansucrase to produce levan, and
PT reacting levan solution in the presence of levan fructotransferase to
PT produce DFA IV -
XX
PS Claim 4; Page 48; 72pp; English.
XX
CC The present sequence is that of *Arthrobacter ureafaciens* K2032 DNA
CC encoding levan fructotransferase (see AAB82301). The DNA was
CC obtained by PCR amplification of genomic DNA using degenerate
CC primers (see AAF30921-22). Expression vector pUDFA81 carrying
CC the levan fructotransferase gene is claimed. Also claimed is
CC *Escherichia coli* JUD81 (K12C 0877BP), prepared by transforming
CC *E. coli* DH5-alpha with pUDFA81. A claimed process for producing
CC difructose dianhydride IV (DFA IV) from sucrose comprises
CC subjecting sugar solution to reaction at room temperature or lower
CC in acidic buffer of pH 3.0-7.0 in the presence of a levansucrase
CC derived from *Zymomonas mobilis* to produce levan, purifying the
CC levan from the reaction solution, and subjecting it to reaction at
CC 25-30 degree C for 3-10 hours in acidic buffer of pH 3.0-7.0 in
CC the presence of levan fructotransferase, preferably obtained by
CC cultivation of *E. coli* JUD81. The product is useful as a
CC low-calorie sweetener.
XX
SQ Sequence 2001 BP; 337 A; 813 C; 588 G; 263 T; 0 other;

Query Match 100.0%; Score 2001; DB 22; Length 2001;
Best Local Similarity 100.0%; Pred. No. 2e-310;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGGTGACCCCGACCTCCCTCGAGACACCCCTGCTCCCTACCGGCGGACCGCCCGGCC 60
DB 1 GCGGTGACCCCGACCTCCCTCGAGACACCCCTGCTCCCTACCGGCGGACCGCCCGGCC 60
QY 61 CGACTGCTCCTCAGCCTAGACGGGCGCCCTCTCTGAGGCTTCTGTCGGGACGGTGAGGCG 120
DB 61 CGACTGCTCCTCAGCCTAGACGGGCGCCCTCTCTGAGGCTTCTGTCGGGACGGTGAGGCG 120
QY 121 ACTGCTCGAACCCTGGTCTCTCTGCGGCGCGCGGTGTGACCGGAGGCTCGAGACGGCA 180
DB 121 ACTGCTCGAACCCTGGTCTCTCTGCGGCGCGCGGTGTGACCGGAGGCTCGAGACGGCA 180
QY 181 CGGCCAGGACCGGTGACGTGACCGGATGACGTGAGGCGCCGACGGATGCTGACGCC 240
DB 181 CGGCCAGGACCGGTGACGTGACCGGATGACGTGAGGCGCCGACGGATGCTGACGCC 240
QY 241 CCTGAACCTGCGCGGCTTCTGGGCTGACGAGCGCTCCACCGCCGACAGCTCTCTCTTAC 300
DB 241 CCTGAACCTGCGCGGCTTCTGGGCTGACGAGCGCTCCACCGCCGACAGCTCTCTCTTAC 300
QY 301 CGCTGCCGGAACAGGGTGGAGCTTCTGCGCGCCACCCGCTCCACGAGAGAACAGCA 360
DB 301 CGCTGCCGGAACAGGGTGGAGCTTCTGCGCGCCACCCGCTCCACGAGAGAACAGCA 360
QY 361 ATGACCGCGGCGATCTACAGCGCGCGCTGCTCAGGAGGCGGCGCGGAGCACTCGCC 420
DB 361 ATGACCGCGGCGATCTACAGCGCGCGCTGCTCAGGAGGCGGCGCGGAGCACTCGCC 420
QY 421 CTGATCTTCGGGGTGTGTCGCGCTCAGCGGGGATCGGCTCCGGGCTCGCTCCGT 480
DB 421 CTGATCTTCGGGGTGTGTCGCGCTCAGCGGGGATCGGCTCCGGGCTCGCTCCGT 480
QY 481 GCGCTTACCAATGACGCCCGCCCGAGCGGCTGGCTCTCGACCCCGCCCGGCTACC 540
DB 481 GCGCTTACCAATGACGCCCGCCCGAGCGGCTGGCTCTCGACCCCGCCCGGCTACC 540
QY 541 ACCACGGCGGCTACCACTGCTACTGCTCCGACAGACAGACCGGCGCGCGGCC 600

DB 541 ACCACGGCGGCTACCACTGCTACTGCTCCGACAGAACAGGCGCGCGGCC 600
QY 601 TGGGACACCGAGACGACGCGCTCGCTTTCAGCGACACCGGACCGTGTATGCGG 660
DB 601 TGGGACACCGAGACGACGCGCTCGCTTTCAGCGACACCGGACCGTGTATGCGG 660
QY 661 CTGCGGCGGCTTCCCGGTGCTGCTGCGGCTGCTGCGGACCGGACCGCA 720
DB 661 CTGCGGCGGCTTCCCGGTGCTGCTGCGGCTGCTGCGGACCGGACCGCA 720
QY 721 GGGTTCGGCGCGGCGGCTGCTGCTGCGGACCGGACCGGACCGGCTGCGCAAG 780
DB 721 GGGTTCGGCGCGGCGGCTGCTGCTGCGGACCGGACCGGACCGGCTGCGCAAG 780
QY 781 TACCAGGACGAGTACTCTACTGCTGACCGGCGGCTTTCAGCTTTCACCGCTGCGC 840
DB 781 TACCAGGACGAGTACTCTACTGCTGACCGGCGGCTTTCAGCTTTCACCGCTGCGC 840
QY 841 GACCCGCTCATGCTCAACACCGGCTGCGGCGGACCGGCGGAGATCGAGAAC 900
DB 841 GACCCGCTCATGCTCAACACCGGCTGCGGCGGACCGGCGGAGATCGAGAAC 900
QY 901 GCGGAGTGTTCGCGACCGGACCGGACCGGCTGCGGACCGGCGGAGATGCTGCG 960
DB 901 GCGGAGTGTTCGCGACCGGACCGGACCGGCTGCGGACCGGCGGAGATGCTGCG 960
QY 961 GTCACTGCGGACGACTGCGGTGACGCGGCTTCTACACCTGCGGACCGGCTGCGGAC 1020
DB 961 GTCACTGCGGACGACTGCGGTGACGCGGCTTCTACACCTGCGGACCGGCTGCGGAC 1020
QY 1021 CTTCCGCGCAACTTCGACTACCGGACCGGCTTTCAGCTTTCAGCTTTCACCGCTGCGC 1080
DB 1021 CTTCCGCGCAACTTCGACTACCGGACCGGCTTTCAGCTTTCAGCTTTCACCGCTGCGC 1080
QY 1081 TTGAGATCACCGGACGAGCGGACCGGCTGCGGACCGGCTGCGGACCGGCTGCGGAC 1140
DB 1081 TTGAGATCACCGGACGAGCGGACCGGCTGCGGACCGGCTGCGGACCGGCTGCGGAC 1140
QY 1141 TACGGCATCGGCTTCCCGCATGACGCTTCTGAGGACCGGCTTTCAGCTTTCAGCTTTCAC 1200
DB 1141 TACGGCATCGGCTTCCCGCATGACGCTTCTGAGGACCGGCTTTCAGCTTTCAGCTTTCAC 1200
QY 1201 TCCACGCGGAGGACTCAGCGCGGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1260
DB 1201 TCCACGCGGAGGACTCAGCGCGGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1260
QY 1261 GTCACCTGCGCATGATGACGCGCGGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1320
DB 1261 GTCACCTGCGCATGATGACGCGCGGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1320
QY 1321 AACTGGAAGTACGCGCGGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1380
DB 1321 AACTGGAAGTACGCGCGGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1380
QY 1381 AACTGATGCTCGCGGAGCTGCGGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1440
DB 1381 AACTGATGCTCGCGGAGCTGCGGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1440
QY 1441 AGCAGCCCGTGGGAGCGCTGACGAGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1500
DB 1441 AGCAGCCCGTGGGAGCGCTGACGAGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1500
QY 1501 ACCGTCGAGCGGCGGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1560
DB 1501 ACCGTCGAGCGGCGGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1560
QY 1561 ATCGGCTGGGACACCGGCGGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1620
DB 1561 ATCGGCTGGGACACCGGCGGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1620
QY 1621 CGGACACGAGCATCGGCAAGTACGAGGACGAGCTTTCAGCTTTCAGCTTTCAGCTTTCAG 1680

Db 1621 CGGCACAGAAATCGGCAAGTACGAGGACAGACCTGTACGTGACCGGAGGACCTCCGAC 1680
Qy 1681 CTCGCGGGTACTCGTCGCGCCCTACTTCGCGAGCGCGCCGCCATCGACCCGCGGCC 1740
Db 1681 CTCGCGGGTACTCGTCGCGCCCTACTTCGCGAGCGCGCCGCCATCGACCCGCGGCC 1740
Qy 1741 CGATCCGTGACCTCGCGCATCTCGTCGACACCCAGAGCGTCGAGTCTTCGTCACGCGC 1800
Db 1741 CGATCCGTGACCTCGCGCATCTCGTCGACACCCAGAGCGTCGAGTCTTCGTCACGCGC 1800
Qy 1801 GGCACACCGTCTCCACAGGTCCTCGTCGACACCCAGAGCGTCGAGTCTTCGTCACGCGC 1860
Db 1801 GGCACACCGTCTCCACAGGTCCTCGTCGACACCCAGAGCGTCGAGTCTTCGTCACGCGC 1860
Qy 1861 TACACGACGCGCGCGCACACTTCACCGGCATCGTCGCGAGATTGGCCAGGCG 1920
Db 1861 TACACGACGCGCGCGCACACTTCACCGGCATCGTCGCGAGATTGGCCAGGCG 1920
Qy 1921 ATCTAGGCGATGACACACACCTCACCAGGCGCGCGCGCGCGCGCGCGCGCGCGAC 1980
Db 1921 ATCTAGGCGATGACACACACCTCACCAGGCGCGCGCGCGCGCGCGCGCGCGCGAC 1980
Qy 1981 AATCGACACGTCCTCGTGGTT 2001
Db 1981 AATCGACACGTCCTCGTGGTT 2001
RESULT 2
AAF30919
ID AAF30919 standard; DNA; 2000 BP.
XX AC AAF30919;
XX
XX 09-JUL-2001 (first entry)
XX Arthrobacter ureafaciens levan fructotransferase DNA.
XX
XX Levan fructotransferase; difructose dianhydride IV; sweetener; ds.
XX OS Arthrobacter ureafaciens.
XX
XX Key Location/Qualifiers
XX CDS 360..1925
XX sig_peptide /*tag= a
XX mat_peptide /*tag= b
XX /*tag= b
XX W0200129185-A1.
XX
XX 26-APR-2001.
XX
XX 19-OCT-2000; 2000WO-KR01183.
XX
XX 19-OCT-1999; 99KR-0045302.
XX
XX (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
XX (REAL-) REALBIOTECH LTD.
XX
XX Rhee S, Song K, Kim C, Ryu E, Lee Y;
XX WPI; 2001-308483/32.
XX
XX Producing difructose dianhydride IV from sucrose, involves reacting
XX sugar solution in the presence of levansucrase to produce levan, and
XX reacting levan solution in the presence of levan fructotransferase to
XX produce DFA IV -
XX
XX Disclosure; Page 8-9; 72pp; English.
XX
XX The present sequence is that of Arthrobacter ureafaciens K2032 DNA
XX encoding levan fructotransferase (see AAB82301). The DNA was

CC obtained by PCR amplification of genomic DNA using degenerate
CC primers (see AAF30921-22). Expression vector pUDFA81 carrying
CC the levan fructotransferase gene is claimed. Also claimed is
CC Escherichia coli JUD81 (KCTC 0877BP), prepared by transforming
CC E. coli DH5-alpha with pUDFA81. A claimed process for producing
CC difructose dianhydride IV (DFA IV) from sucrose comprises
CC subjecting sugar solution to reaction at room temperature or lower
CC in acidic buffer of pH 3.0-7.0 in the presence of a levansucrase
CC derived from Zymomonas mobilis to produce levan, purifying the
CC levan from the reaction solution, and subjecting it to reaction at
CC 25-30 degree C for 3-10 hours in acidic buffer of pH 3.0-7.0 in
CC the presence of levan fructotransferase, preferably obtained by
CC cultivation of E. coli JUD81. The product is useful as a
CC low-calorie sweetener.
XX
XX Sequence 2000 BP; 337 A; 813 C; 587 G; 263 T; 0 other;
SQ
Query Match 100.0%; Score 2000; DB 22; Length 2000;
Best Local Similarity 100.0%; Pred. No. 2.9e-310;
Matches 2000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 CGGTGACCGCGACTTCCTCGAGACACCGTCCCTACCGCGCGACCGCGCGCC 61
Db 1 CGGTGACCGCGACTTCCTCGAGACACCGTCCCTACCGCGCGACCGCGCGCC 60
Qy 62 GACTGCTCTCAGCTTAGAGCGGCCCTCTCTCGAGGTCCTTCTCGGGGAGGTGAGCGA 121
Db 61 GACTGCTCTCAGCTTAGAGCGGCCCTCTCTCGAGGTCCTTCTCGGGGAGGTGAGCGA 120
Qy 122 CTGCGTCGAACCTGGTCTCTCTGGGGCGCGCGGTGTACCGCGAGCTCGAGACGCGAC 181
Db 121 CTGCGTCGAACCTGGTCTCTCTGGGGCGCGCGGTGTACCGCGAGCTCGAGACGCGAC 180
Qy 182 GCGCAGGAACCGTGACGTCACCGCGATGACGTCGAGGCGCGCGAGTGTCTGACGCC 241
Db 181 GCGCAGGAACCGTGACGTCACCGCGATGACGTCGAGGCGCGCGAGTGTCTGACGCC 240
Qy 242 CTGAACCTGCGCGTTCCTGGGTGACGAGCGTCCACCGCGCGAGCTCTCTCTTACC 301
Db 241 CTGAACCTGCGCGTTCCTGGGTGACGAGCGTCCACCGCGCGAGCTCTCTCTTACC 300
Qy 302 GCTGCCGGAACAGGTGGAGCTTCGTGGCGGCCACCGCTCCACGAGAGAAACAGCAA 361
Db 301 GCTGCCGGAACAGGTGGAGCTTCGTGGCGGCCACCGCTCCACGAGAGAAACAGCAA 360
Qy 362 TGACGCGCGCATCTCAGCGCGCGCTGTCTCCAGGAGCGCGCGCGAGACTCGGCC 421
Db 361 TGACGCGCGCATCTCAGCGCGCGCTGTCTCCAGGAGCGCGCGCGAGACTCGGCC 420
Qy 422 TGATCTTTCGGGCGTGTGTGCGGCTTCAGCGCGCGGATCCGCTCCGGGTCGCTGCGTG 481
Db 421 TGATCTTTCGGGCGTGTGTGCGGCTTCAGCGCGCGGATCCGCTCCGGGTCGCTGCGTG 480
Qy 482 CCGTCTACACATGACCGCGCGCGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 541
Db 481 CCGTCTACACATGACCGCGCGCGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 540
Qy 542 CCCACGCGCTACAGCTGTACTACTCTCCAGTCCGACAGAAACAGCGCGCGCGCGCT 601
Db 541 CCCACGCGCTACAGCTGTACTACTCTCCAGTCCGACAGAAACAGCGCGCGCGCGCT 600
Qy 602 GGGACACGCGAGCAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 661
Db 601 GGGACACGCGAGCAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
Qy 662 TGGGCGCGGACTTCCCGTGTGGTCCGGGTGCGGGTGTGCGGCGCGCGCGCGCGCGCG 721
Db 661 TGGGCGCGGACTTCCCGTGTGGTCCGGGTGCGGGTGTGCGGCGCGCGCGCGCGCGCG 720
Qy 722 GGTTCGGCGCGCGCGGTGCTGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 781
Db 721 GGTTCGGCGCGCGCGGTGCTGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780

QY 782 ACCAGGAGCAGTACCTCTACTGGTCCAGCAGCGGGGGTTACGTTACCGCGCTGCCCG 841
Db 781 ACCAGGAGCAGTACCTCTACTGGTCCAGCAGCGGGGGTTACGTTACCGCGCTGCCCG 840
QY 842 ACCCGCTCATCGTCAACACCGAGCGTCCGCGGCCACACGCGCGCGGAGATCGAGAACG 901
Db 841 ACCCGCTCATCGTCAACACCGAGCGTCCGCGGCCACACGCGCGCGGAGATCGAGAACG 900
QY 902 CCGAGTGGTTCGCGGACCCCAAGATCCACTGGGACACCCCGCGGAGAAATGGTCTGGG 961
Db 901 CCGAGTGGTTCGCGGACCCCAAGATCCACTGGGACACCCCGCGGAGAAATGGTCTGGG 960
QY 962 TCATCGGACGACTCGCGGTACGCGGCTTCTACACCTCGCGGAACTCGCGACTGGACAC 1021
Db 961 TCATCGGACGACTCGCGGTACGCGGCTTCTACACCTCGCGGAACTCGCGACTGGACAC 1020
QY 1022 TTCGCGCGCAACTTCGAGTACCGAACCGACCGCTCGCGGCGCATCGAGTCCCGGACCTGT 1081
Db 1021 TTCGCGCGCAACTTCGAGTACCGAACCGACCGCTCGCGGCGCATCGAGTCCCGGACCTGT 1080
QY 1082 TCGAGATCACCGCAGACGAGCGGACACGCTACTGGGTGCTCGCGCGCAGCATGGAGCGCT 1141
Db 1081 TCGAGATCACCGCAGACGAGCGGACACGCTACTGGGTGCTCGCGCGCAGCATGGAGCGCT 1140
QY 1142 AGGGCATCGCGCTCCCGCATGAGTACGCTACTGGACAGGCACTGGGAGCGCGAGCACT 1201
Db 1141 AGGGCATCGCGCTCCCGCATGAGTACGCTACTGGACAGGCACTGGGAGCGCGAGCACT 1200
QY 1202 TCACGCGGACGACCTTCACCCCGCAATGCTCGACTGGGGCTGGGACTGGTACGCGGGCG 1261
Db 1201 TCACGCGGACGACCTTCACCCCGCAATGCTCGACTGGGGCTGGGACTGGTACGCGGGCG 1260
QY 1262 TCACCTGGCCATCGATCGACGCGCGGAGACCAAGCGCTCGGCCATCGCGTGGATGAACA 1321
Db 1261 TCACCTGGCCATCGATCGACGCGCGGAGACCAAGCGCTCGGCCATCGCGTGGATGAACA 1320
QY 1322 ACTGGAAGTACCGCGACGCGAGCTGCCACCGAGCGATTCGACGGGCTACACGGGCGAGA 1381
Db 1321 ACTGGAAGTACCGCGACGCGAGCTGCCACCGAGCGATTCGACGGGCTACACGGGCGAGA 1380
QY 1382 ACTGATCTCGCGGAGCTGCGGCTCGCCGAGACGCTGGCGGCTGGTACACCTCTCTGA 1441
Db 1381 ACTGATCTCGCGGAGCTGCGGCTCGCCGAGACGCTGGCGGCTGGTACACCTCTCTGA 1440
QY 1442 GCACCCCGTGGCAGCGGTGACGAATACTAGTCAACCGCCACACCACTCCCGGACCGGA 1501
Db 1441 GCACCCCGTGGCAGCGGTGACGAATACTAGTCAACCGCCACACCACTCCCGGACCGGA 1500
QY 1502 CCGTCGAGCGGACGCGCTCTGCGCATGGAGCGGCGCATACGAGATCGAGCTCGACA 1561
Db 1501 CCGTCGAGCGGACGCGCTCTGCGCATGGAGCGGCGCATACGAGATCGAGCTCGACA 1560
QY 1562 TCGCCTGGGACACCGCGACGACGTCGGCATCTCGGTGGCGGCTCCCGCGACGGAACCC 1621
Db 1561 TCGCCTGGGACACCGCGACGACGTCGGCATCTCGGTGGCGGCTCCCGCGACGGAACCC 1620
QY 1622 GGCACAGCAATCGGCAAGTACGAGCAGACCTTACGTCGACCGAGACCTCCCGACC 1681
Db 1621 GGCACAGCAATCGGCAAGTACGAGCAGACCTTACGTCGACCGAGACCTCCCGACC 1680
QY 1682 TCGCGGGGTACTCGCTCGCGCCCTACTCGCGAGCGCGCGCCCATCGACCCCGCGGCC 1741
Db 1681 TCGCGGGGTACTCGCTCGCGCCCTACTCGCGAGCGCGCGCCCATCGACCCCGCGGCC 1740
QY 1742 GATCGGTGACCTCGCGCATCTCTCGTCGACACCCAGAGCGTCTGAGTCTTCGTCACAGCGG 1801
Db 1741 GATCGGTGACCTCGCGCATCTCTCGTCGACACCCAGAGCGTCTGAGTCTTCGTCACAGCGG 1800
QY 1802 GCCACACGCTGCTCTCCAGCAGGTCCACTTCGCGGAGGGGACACGGAATCTCGCTCT 1861
Db 1801 GCCACACGCTGCTCTCCAGCAGGTCCACTTCGCGGAGGGGACACGGAATCTCGCTCT 1860
QY 1862 ACACCGCGGGGCGCGCCGACACATTCACCGGCGATCGTCTCGCGGAGATTGGCCAGGGA 1921

Db 1861 ACACCGACGGCGCGCCGACACTTACCGGCGATCGTCTCGGAGATTGGCCAGGCCA 1920
QY 1922 TCTAGGCGATGACACACACCGCTTACCGAAGCGCGCGCCGAGAGACGAGCGCGGACA 1981
Db 1921 TCTAGGCGATGACACACACCGCTTACCGAAGCGCGCGCCGAGAGACGAGCGCGGACA 1980
QY 1982 ATCGACACGCTCTCGTCTGTT 2001
Db 1981 ATCGACACGCTCTCGTCTGTT 2000

RESULT 3
ABL54835
ID ABL54835 standard; DNA; 1752 BP.
AC ABL54835;
XX
DT 01-JUL-2002 (first entry)
XX
DE Levan fructotransferase related DNA sequence.
XX
KW Levan fructotransferase; LFTase; difructose dianhydride; DFA IV;
KW levan; gene; ds.
XX
OS Arthrobacter nicotinovorans.
XX
PH Key Location/Qualifiers
FT CDS 4..1461
FT /*tag= a
FT /product= "levan fructotransferase related protein"
XX
PN JP2002017366-A.
XX
PD 22-JAN-2002.
XX
PF 06-JUL-2000; 2000JP-0205756.
XX
PR 06-JUL-2000; 2000JP-0205756.
XX
PA (NIPT) NIPPON TENSU SEITO KK.
PA (KAGA-) KAGAKU GIUTSU SHINKO JIGYODAN.
XX
DR WPI; 2002-287313/33.
XX
XX A new levan fructotransferase -
PS Example 1; Fig 4; 15pp; Japanese.
XX
CC The invention relates to a new levan fructotransferase (LFTase). LFTase
CC decomposes the beta-2,6-fructoside bond of levan in polyfructan to
CC produce difructose dianhydride IV (DFA IV). The LFTase of the invention
CC is used for the preparation of DFA IV. The current sequence represents
CC a levan fructotransferase related DNA sequence.
SQ Sequence 1752 BP; 351 A; 592 C; 506 G; 303 T; 0 other;

Query Match 47.2%; Score 945.2; DB 24; Length 1752;
Best Local Similarity 74.3%; Pred. No. 2.9e-142;
Matches 1237; Conservative 0; Mismatches 413; Indels 14; Gaps 3;

QY 314 AGGTGGAGCTTCTCGCGCCACCCCGTCCACGAGAGAAC--AGCAATGACCGCGC 371
Db 42 ACGTTTGGCGTGTCTCGCGCGCGCGCAACACTGAGAGGAACGAATCGATGACGTATGA 101
QY 372 CATCTACCGCGCGCGTGTCTCCAGGAGCGCGCGGACACTCGCCCTGATCTTCGG 431
Db 102 CATCTCTCCCGCACTGCCCTGCAAGGTGACAGGGTGTGGTCTGGACATTTTCATGAG 161
QY 432 CGGTGCTGTGCGCTGCAAGCGCGGCGCATCGCTCCGGGCTCGCTCCGCGCTACCA 491
Db 162 CAATGCCATTCCTCGTGGCGCGCCCA-----CGCCAGGACATCCCTCCGGGCGATCTACCA 215

```
Qy 492 CATGACGCCCCCAGCGCTGGCTCTCGACACCCCAACGCCCGGTCACACCCACGGCGC 551
Dy 216 CATGACCCCGCCCTCGGCTGCTATGTATCGCAGCAGCCCGTACATACAAACGGCGC 275
Qy 552 CTACCAGCTGTACTACCTCGACTCGGACCAAGAACAGCCCGCGCTGGGACCGC 611
Dy 276 CTACAGCTGTACTACCTCGACTCGGACCAAGAACAGCCCGCGGATGGGACCGC 335
Qy 612 GAGCACGACGACGCGCTCGCTTTCACGACACCGGACCGCTGATCGCGCTCGGCGCGA 671
Dy 336 GACCACGCGCAGGAGTGTCTTACACCCACCATGGAGTGTGATGCCAATGCAACCGA 395
Qy 672 CTTCCCGCTGTGTCGCGGTGCGGCGTGTGTCGACCGGACGAGGAGGTTGCGGCG 731
Dy 396 CTTCCCGCTGTGTCGCGGTGCGGCGTGTGTCGACCGGACGAGGAGGTTGCGGCG 455
Qy 732 CGGCGCGTGTGTCGCGGTGCGGCGTGTGTCGACCGGACGAGGCGTGTGTCGAGGAGCA 791
Dy 456 CGGCGAGTGTGTCGCGGTGCGGCGTGTGTCGACCGGACGAGGCGTGTGTCGAGGAGCA 509
Qy 792 GTACCTCTACTGTCGACGCGGCGGTTCACGCTTACCGCCCTCGGCGGACCGCGTCAT 851
Dy 510 GTACCTTTACTGTCGACGCGGCGGTTCACGCTTACCGCCCTCGGCGGACCGCGTCAT 569
Qy 852 CGTCAACACGAGCGTGTGCGGCGTGTGTCGACCGGACGAGGAGGTTGCGGCGT 911
Dy 570 TGTGAACACTGTGACGAGCGGCGTGTGTCGACCGGACGAGGAGGTTGCGGCGT 629
Qy 912 CCGGACCCCAAGATCCACTCGGACGCGGCGGCGGAGAGTGTGTCGCTCATCGGACG 971
Dy 630 CCGGACCCCAAGATCCACTCGGACGCGGCGGCGGAGAGTGTGTCGCTCATCGGACG 689
Qy 972 ACTCGGTACGCGCGGTGTCTACACTCGCGGACCTCGGCGGACGAGTGTGTCGCGGCA 1031
Dy 690 GCGCGCTACGCTGCTCTACACTCTCCCAACCTCGGCGGATGGCAATGGAAGTCCAA 749
Qy 1032 CTTGACTACCCGACACGCGCTCGGCGGATCAGTGTGCGGCGGAGTGTGCGGAGTAC 1091
Dy 750 CTTGACTACCCGACACGCGCTCGGCGGATCAGTGTGCGGCGGAGTGTGCGGAGTAC 809
Qy 1092 CGCAGACGCGGACACGCGCTCGGCGGATCAGTGTGCGGCGGAGTGTGCGGAGTAC 1151
Dy 810 CGCAGGACGCGGACACGCGCTCGGCGGATCAGTGTGCGGCGGAGTGTGCGGAGTAC 869
Qy 1152 CTTCCCGATGACGCTGCTGACGAGGACCTCGGCGGAGTGTGCGGAGTGTGCGGAGTAC 1211
Dy 870 CTTGCCATGACCTTGGCTTACTGACAGGTTTATGGAAGGACGAGTGTGCGGAGTAC 929
Qy 1212 CGACTACCCCGGATGCTGACTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1271
Dy 930 CAACCTCACACACAGTGGCTGACTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 989
Qy 1272 ATCGATCGACGCGCGGAGACCAAGCGCTCGGCGGATGAGCAACTGGAAGTA 1331
Dy 990 GCGCGTGAAGACCTGAGACCAAGCGCTTGGCAGAGGCTGGGATGAGCAACTGGAAGTA 1049
Qy 1332 CGCGGACGCGGAGTGTGCGGCGGATGAGCAACTGAGCAACTGAGCAACTGAGCAACT 1391
Dy 1050 TGCGCGCGGCAACGTGCGGCGGAGTGTGCGGCGGATGAGCAACTGAGCAACTGAGCA 1109
Qy 1392 CCGGAGTGTGCGGCGGAGTGTGCGGCGGATGAGCAACTGAGCAACTGAGCAACTGAGCA 1451
Dy 1110 GCGGAGTGTGCGGCGGAGTGTGCGGCGGATGAGCAACTGAGCAACTGAGCAACTGAGCA 1169
Qy 1452 GCGGAGTGTGCGGCGGAGTGTGCGGCGGATGAGCAACTGAGCAACTGAGCAACTGAGCA 1511
Dy 1170 TCGGCGGCTTTCGAACTATGCACTTCCAGCACTTCCAGCACTTCCAGCACTTCCAGCA 1229
Qy 1512 CAGCGCGTCTGCGATGAGCACTGAGCACTGAGCACTGAGCACTGAGCACTGAGCACTGAGCA 1571
Dy 1230 CAGTTTGTGACTTCCGTGGGCGGCGGCTGATGAACTGCACTTCCAGCACTTCCAGCA 1289
Qy 1572 CACCGGACGAACTGCGGCTGCGGTGGCGGCTGCGGCGGAGTGTGCGGCGGAGTGTGCGG 1631
```

```
Dy 1290 CACGCGACGAACTGGGAGTCTCGGTGGCGGCTCGGTCCGATGCCAGCCCATACGAA 1349
Qy 1632 CATCGGAAGTACGAGCAGACCTGTAGTTCGACGCGGAGGACCTCCGACCTCGCGGGA 1691
Dy 1350 CATCGGAAGTACGAGTTCGAGTGTAGTTCGATCGGCGATCTTCGGAGCAACGCGTTA 1409
Qy 1692 CTCGCTGCGCCCTACTCTCGGAGCGCGCGCCCATCGACCCCGCGCGCCCGATCCGTGCA 1751
Dy 1410 TGCGCTGCGACCTTACACCCCGCGCGCCCATCGATCGAAGCCAGATCCGTCCA 1469
Qy 1752 CTTGCGATCTCTGTCGACACCCAGAGTGTGAGTCTTCTCAACGCGCGCGGACACCGT 1811
Dy 1470 CTTGCGATCTTCTAGACACCCAAAGTGTGAGTGTGTTGTAATTTCCGCGCACCGT 1529
Qy 1812 GCTCTCCAGCAGTCCACTTCGCGGAGGCGACACGGAATCTCGTCTACCGGACG 1871
Dy 1530 GTTTCGAGCAGTGTACCTTCGCGGCGCGGAGACCGGGATCTCCCTCTATCGGAGCG 1589
Qy 1872 CGGCGCGCACACTTCACCGGATCTGTCGCGGAGATTTGGCGGAGTCTTAGGCGAT 1931
Dy 1590 CGGTCCGCGCAACTTCACCGGATCACCATCCGCGAGTTCGGGAACCCCATCTAAGCGT 1649
Qy 1932 GCACACACACCGTCTACCGAAGCGCGCGCGCGCGGAGACGAGCG 1975
Dy 1650 CGTCCACGCGGTGGAAGGACGCGGACGCTGCAGCAGCGG 1693
```

RESULT 4

```
AA119827
ID AAX19827 standard; DNA; 1551 BP.
XX
AC AAX19827;
XX
DT 10-JUN-1999 (first entry)
DE
DE Arthrobacter nicotinovorans levan fructotransferase encoding DNA #2.
XX
XX Arthrobacter nicotinovorans; levan fructotransferase; ds.
XX
XX Arthrobacter nicotinovorans.
XX
XX JPL1069978-A.
XX
PD 16-MAR-1999.
XX
PF 28-AUG-1997; 97JP-0232421.
XX
PR 28-AUG-1997; 97JP-0232421.
XX
PA (NIOC ) NIPPON OIL CO LTD.
XX
XX WPI; 1999-247463/21.
XX
XX P-PSDB; AAY04105.
```

Levan fructotransferase gene - for recombinant production of levan fructotransferase

Claim 4; Page 10-11; 14pp; Japanese.

The present sequence encodes Arthrobacter nicotinovorans levan fructotransferase. The present invention also describes a method for the preparation of levan fructotransferase in which a transformant is cultured in a medium and levan fructotransferase is collected from the culture. The method can prepare levan fructotransferase in a large amount.

Sequence 1551 BP; 315 A; 524 C; 443 G; 269 T; 0 other;

Query Match 46.6%; Score 931.8; DB 20; Length 1551;
Best Local Similarity 75.8%; Pred. No. 4e-140;
Matches 1184; Conservative 0; Mismatches 367; Indels 12; Gaps 2;

```
QY 361 ATGACGCCGGCATCTACGCGCGCGCGTGTCTCCAGGAGCGCGCGCGGACGACCTCGCC 420
Db 1 ATGACGATGACATCTCTCGCGCACTGCGCTGCAAGGTGCAAGGTGCTGCTTGGCA 60
QY 421 CTGATCTTCGGCGGTGCTGTGCGGCTGACGCGCGGATCGCTCCGGCTCGCTCCGT 480
Db 61 CTCTTTCATGAGCAATGCAATCTCCGTTGCGCGCCCA-----CGCCAGGACATCCCTCCGG 114
QY 481 GCGCTCTACACATACGCGCGCGCGTGTGCTGCGACCGCCCAACCGCGGTCAACC 540
Db 115 GCGATCTACACATGACCGCGCGCTCGGCTGTGATGTGATCGCGGACCGGTACAT 174
QY 541 ACCACGCGCTACAGCTGTACTACTGTGACCTCCGACGAGAACAAACGCGCGCGCG 600
Db 175 ACAACGCGCGCTACAGCTGTACTACTGTGACCTCCGCGGACGAGAACAAACGCGCGCG 234
QY 601 TGGGACCGGACGACGACGCGCGCTGCTTCACGACCGGACCGGACCGGTGATGCGG 560
Db 235 TGGGACCGGACGACGACGCGCGCTGCTTCACGACCGGACCGGACCGGTGATGCGG 294
QY 661 CTGCGCGCGGACTTCCCGCTGTGCTCGGCTCGGCGTCTGCGGACCGGACCGGACCG 720
Db 295 ATGCAACCGGACTTCCCGCTGTGCTCGGATCGGACGAGTAGTGAGTGGATGCGCA 354
QY 721 GGGTTCGGCGCGCGGCTGTGCTGCGCTGCGACCGGACCGGACCGGCTCGCGAAG 780
Db 355 GGGTTCGGCGCGCGGCTGTGCTGCGCTGCGACCGGACCGGACCGGACCGGACCG 408
QY 781 TACGAGGACGACTACTCTACTGTGCTGACGCGGCGGCTTCACGTTTCACCGCGCTGCC 840
Db 409 TTCCAGGAACAGTACTCTTACTGTGCTGCGGATGCGGCTACTCTTTCACCGCATTGCT 468
QY 841 GACCCGCTCATCTGCTCAACACCGACGCTGCGCGCGGACCGGACCGGACCGGACCG 900
Db 469 GACCCGCTCATCTGCTCAACACCGACGCTGCGCGCGGACCGGACCGGACCGGACCG 528
QY 901 GCGGAGTGGTTCGGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 960
Db 529 GCAGAAATGTTCCGGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 588
QY 961 GTCATCGGACGACTCGGTGACGCGCGCTTACACTCTGCGGACCGGACCGGACCGG 1020
Db 589 GTCATCGGACGCGGCTGACGCTGCTTACACTCTGCGGACCGGACCGGACCGGACCG 648
QY 1021 CTTCGCGCACTTCGACTACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 1080
Db 649 TGAAGTCCAACTTCGACTACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 708
QY 1081 TTCGAGATCACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 1140
Db 709 TTCGAAATGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 768
QY 1141 TACGCACTCGGCTCCCATGACGCTACGCTACTGACGAGGACCGGACCGGACCGGACCG 1200
Db 769 TACGCACTCGGCTCCCATGACGCTTTCGCTACTGACGAGGTTTCATGAAACCGGACCG 828
QY 1201 TTCGACGCGGACGACTCACCGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 1260
Db 829 TTCATCGCGGACCACTCACACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 888
QY 1261 GTCACCTGGCCATCATGATGACGCGCGGACCGGACCGGACCGGACCGGACCGGACCG 1320
Db 889 GTGACCTGGCGCGGCTGGAAGCACTGAGACCAAGCGGCTTTCGACGAGCGGATGAAC 948
QY 1321 AACTGGAAGTACGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 1380
Db 949 AACTGGAATATGCGCGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 1008
QY 1381 AACTCGATGTCGCGGAGCTCGGCTCGCGGACCGGACCGGACCGGACCGGACCGGACCG 1440
Db 1009 AATTCCATCACGCGGAGCTCAGGCTCGAGCGGCAATCGGCGGCTGTTACACCTTGCTC 1068
QY 1441 AGCACCCCGGTGCGAGCGCTGACGAACTACGTACCGCGGACCGGACCGGACCGGACCGG 1500
```

```
Db 1069 AGCAGCGCGGTTCGCGCGCTTCGAACTATGCCACTATGCCACTTCAGCACCCCTTCGCGACCGC 1128
QY 1501 ACCGTTCGAGCGGACGCGCTCTGCTGCAATGGAACGAGCGGACATAGAGATCGAGTTCGAC 1560
Db 1129 ACAGTCAACGCGGAGTTTCGTAATTCGCTGAGCGCGCGGCTGATGAACTGGAATCTGAT 1188
QY 1561 ATCGCTGGGACACGCGGACGAGCAATCGGCACTCTCGGTGGCGGCTTCCCGGACGGAACC 1620
Db 1189 ATTTTCATGGGACACGCGGACGAGAACTGGAGTCTCGGTGGCGGCTCGTCCGATGGCAGC 1248
QY 1621 CGGACACGAACTCGGCAATGAGTACGAGGACGAGCACTGTAGCTCGAGCGGAGACCTCCGAC 1680
Db 1249 CGCCCATACGAACTCGGCAATGAGTACGAGTTCGAGTTCGATCGATCGCGCATCTCTCGAG 1308
QY 1681 CTCGCGGCTACTCTGCTGCGCGCTTACTCGCGAGCGCGCGGCTTCCCGGACGAGGACCG 1740
Db 1309 CAAAGCGGTATGCGCTGGCAGCTTACACCGCGCGGCTTACGATCGGAGACCGG 1368
QY 1741 CGATCGCTGCACTCGCATCTCTGTCGACACCGGAGCGTCTGAGGTCTTCTGCAACGCC 1800
Db 1369 AGATCGCTCCACTCGCATCTTGTAGACACCGAAAGTGTGAGGTGTTCGTAATTC 1428
QY 1801 GGCACACGCTGCTCTCCAGCAGGTCCACTTCGCGGAGGCGGACACGGAATCTCGCTC 1860
Db 1429 GGGCACACGCTGTTTCGCGAGGTTGCACTTCGCGCGCGGCGGACACGGAATCTCTCTC 1488
QY 1861 TACACGAGCGGCGCGGCGGACGAGTTCACCGGCACTTCGCTCGCGGAGATTCGCGAGGCG 1920
Db 1489 TATCGGAGCGGCGGCTCGGCGCACTTCACCGGCACTTCGCGGAGTTCGCGAGGACCG 1548
QY 1921 ATC 1923
Db 1549 ATC 1551

RESULT 5
ABL54834
ID ABL54834 standard; DNA; 1467 BP.
XX
AC ABL54834;
XX
DT 01-JUL-2002 (first entry)
XX
DE Levan fructotransferase encoding sequence.
XX
KW Levan fructotransferase; LFTase; difructose dianhydride; DFA IV;
KW levan; gene; ds.
XX
OS Arthrobacter nicotinovorans.
FH Key Location/Qualifiers
FT CDS 4..1461
FT /tag= a
FT /product= "levan fructotransferase"
XX
PN JP2002017366-A.
XX
PD 22-JAN-2002.
XX
PF 06-JUL-2000; 2000JP-0205756.
XX
PR 06-JUL-2000; 2000JP-0205756.
XX
PA (NIPT ) NIPPON TENSAI SEITO KK.
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
DR WPI; 2002-287313/33.
DR P-PSDB; ABB10073.
XX
PT A new levan fructotransferase
XX
PS Claim 3; Fig 2; 15pp; Japanese.
```

XX The invention relates to a new levan fructotransferase (LFTase). LFTase
CC decomposes the beta-2,6-fructoside bond of levan in polyfructan to
CC produce difructose dianhydride IV (DFA IV). The LFTase of the invention
CC is used for the preparation of DFA IV. The current sequence represents
CC the levan fructotransferase encoding sequence.
XX
SQ Sequence 1467 BP; 304 A; 494 C; 419 G; 250 T; 0 other;

Query Match 45.6%; Score 912; DB 24; Length 1467;
Best Local Similarity 77.0%; Pred. No. 5,8e-137;
Matches 1126; Conservative 0; Mismatches 330; Indels 6; Gaps 1;

Qy 466 CCGGGCTGCTCGTGGCGGTACACATGACGGCCCCCGGCTGCTGCGACCC 525
Db 7 CAGGCATCCCTCGGGGATACACATGACCCCGCTCGGGCTGGCTATGTATCCG 66

Qy 526 CAACGCCGCTCACACCCAGCGGCTTACCAGTGTACTACTTGCATCCGACAGAAC 585
Db 67 CAGGCATCCCTGACATACAAACGGGCGCTTACCAGTGTACTACTTGCATCCGACAGAAC 126

Qy 586 AACGGCCCGGCGCTGGGACACGCGAGACACGACGCGCTTCCACGACACAC 645
Db 127 AACGGACCGGCGGATGGGACACGCGACACCGCGGAGGTCTTACACCCACAT 186

Qy 646 GGCACCGTGTGCGGCTGCGGCGGCTTCCCGTGTGTCGGGTGCGGCTGCTGCGG 705
Db 187 GGAGTGTGTGATGCCAATGCAACCGGACTTCCCGTGTGTCGGGATGGGATGAG 246

Qy 706 ACCGGACACGCGAGGTTCCGGCGCGGCGGCTGCTGCGGCTGCGGACCGCGAC 765
Db 247 ACCGCCAACACCGGCGGCTTCCGGCGCGGCGGCTGCTGCGGCTGCGGACCGCGAC 306

Qy 766 GACGGCGTCCCAAGTACACGAGCAGTACCTTCTACTGTGTCGACGCGGGTTCACG 825
Db 307 GACG-----GAAATTCAGAGAACAGTACCTTTACTGTGTCACGGATGGCGGTACTCC 360

Qy 826 TTCACCGCCCTGCCGACCGCGTTCATCGTCAACACCGACGCTGCGCGCGACACCGCC 885
Db 361 TTCACCGCATTTGCTGACCGCGCTCATTTGTAACACTGATGACGACGCGCCACCGCC 420

Qy 886 GCCAGATGAGAACCGCGAGTGTTCGGGACCGCCACCGACCTGACCTGGGACACCGCCG 945
Db 421 GCCAGTGGAGACGACGAGATGTTCCGGACCGCGGAGTTCACCTGCGGACGCGACGCG 480

Qy 946 GGAGATGGGTCTCGTCACTCGGACGACTGCGGTACGCGGCTTCTACACCTCGCGGAAC 1005
Db 481 AACAGTGGGTCTGTGTCATCGGACGGCGCGCTACGCTGCTTCTACACCTCTCCCAAC 540

Qy 1006 CTGCGGACTGGACACTTCGCGCACTTCGACTACCGAACCGCCCTCGGGGCGATC 1065
Db 541 CTGCGGATTTGGCAATGGAAGTCCAACTTCGACTACCCCAACCGCCCTCGCGGTATC 600

Qy 1066 GAGTGGCCCGACCTTTCGAGATCACCGACAGCGGACCGGACCTGGGTGCTCGCC 1125
Db 601 GAATGCGCGGATCTGTTGAAATACCGACGAGAGAACCGCGGCTGCTGCTCGG 660

Qy 1126 GCCAGTGGACGCTACGCGGCTGCGCTCCCGATGAGTACGCTTCTGACAGAGGAC 1185
Db 661 GCGAGATGGACGCTTACAGATTCGCTTGGCCATGACCTTTGCTTCTGACAGAGTTCA 720

Qy 1186 TGGACCGGCGAGAGCTTCCACCGCGAGACCTTCCCGCGCAATGGCTGACTGGGGGTGG 1245
Db 721 TGGAACGCGACAGCATTCATCGCGGACAACTTACACACAGTGGCTTGTGAGGGATG 780

Qy 1246 GACTGTGACGGCGGCTGACCTGGCCATCGATCGACGCGGCGCGGACCAACCGCTCGCC 1305
Db 781 GACTGTGACGCGCGGCTGACCTGGCGCGGCTGGAAGCACCTGAGACCAAGCGGCTGCC 840

Qy 1306 ATCCGCTGGATGAACACTGAACTAGCGCGACGCGGCTCCCGACCGGATCCGAC 1365
Db 841 ACAGCGGTGATGAACAACTGGAATAATGCGCGCGCGCAACGTCGCCGCGGCTCGAT 900

Qy 1366 GGCTACACAGCGGACAGAACTCGATCGTCGCGAGGTGGGTCTGCCCGACACGCTTGGGGC 1425
Db 901 GGCTATAACGGGCAAAATTCATCAGCGCGGAGCTCAGGCTCGAGCGCAATTCGGGGC 960

Qy 1426 TGGTACACCTTCCTGAGCACCCCGTGGCAGCGTGGAGAACTAGTCCACCGCCACAC 1485
Db 961 TGGTACACCTTGTCTGAGCACCCCGTTCGCGGCTTTCGAACTATGCCACCTCCAGCAC 1020

Qy 1486 ACATCTCCCGGACCGGACCGTTCGAGCGCGGCTCTCTGCCATGGAAGCGGCGCATAC 1545
Db 1021 ACCCTTCGGGACCGCACAGTCAACGGCAGTTTCGTACTTCCGTGGAGCGCGCGGTAT 1080

Qy 1546 GAGATCGAGCTCGATCGCTGGGACACCGCGAGAACGTCGGCATCTCGTGGGCGCG 1605
Db 1081 GAACTGGAACTCGATATTTTCATGGGACACGCGAGCAAGCTGGGAGTCTCGGTGGCGC 1140

Qy 1606 TCCCGCGAGCAACCGCGCACACAACTCGGCAAGTACGAGCAGACCTCTACGTCGAC 1665
Db 1141 TCGTCCGATGCGCGCCATACGAACTACGCAAACTACGTTGAGGTGTGCTGCTGAT 1200

Qy 1666 CGAGACCTCTCGGCTGCGCGGTTACTCTGCTCGCCCTTACTCGCGAGCGCGCGCC 1725
Db 1201 CGCGCATCTCGAGCAAGCGGTTATGCGCTGGCACCTACACCGCGCGCGCGCC 1260

Qy 1726 ATCGACCGCGCGCGGCTGCGTGCACCTGCGCATCTCTGTCGACACCAAGCTCGAG 1785
Db 1261 ATCGATGCGAAACGCGCATCGCTCCACCTGCGCATCTTTGTAGACACAAAGTCTTGAG 1320

Qy 1786 GTCTTCGTCAACGCGCGCACACGCTGCTCTCCAGCAGGTCCACTTCGCGAGGCGAC 1845
Db 1321 GTGTTCTGTAATTCGCGGACACGCTGTTTCGCGCAGGTGCTCTCGCGCGCGGAC 1380

Qy 1846 ACGGAACTCTCGCTTACACGAGCGCGCGCGCGCACACTTTCACCGGATCTGCTCGC 1905
Db 1381 ACGGGATCTCCCTCTATGCGGACGCGGCTCGGCGCACTTCACCGGATCACCATCCG 1440

Qy 1906 GAGATGGCCAGGCGATCTAGG 1927
Db 1441 GAGTTCGGGAACCGCATCTAAG 1462

RESULT 6
AA19826
ID AAX19826 standard; DNA; 1452 BP.
XX AC AAX19826;
XX
XX 10-JUN-1999 (first entry)
XX
XX Arthrobacter nicotinovorans levan fructotransferase encoding DNA #1.
XX Arthrobacter nicotinovorans; levan fructotransferase; ds.
XX Arthrobacter nicotinovorans.
XX
XX JPI1069978-A.
XX
XX 16-MAR-1999.
XX
XX 28-AUG-1997; 97JP-0232421.
XX
XX 28-AUG-1997; 97JP-0232421.
XX (NIOC) NIPPON OIL CO LTD.
XX
XX WPI; 1999-247463/21.
XX P-PSDB; AAY04104.
XX
XX Levan fructotransferase gene - for recombinant production of levan
XX fructotransferase
XX
XX Claim 2; Page 9-10; 14pp; Japanese.
XX

Qy	1366	GGCTACAACGGCGAGAACTCGATCGTCCGCGAGCTGCGGCTCGCCCGACACAGCTTGGCGGC	1425
Db	895	GGCTATAACGGCGGCAAAATTCATCAGCGCGAGCTCAGGCTCGAGCGCAATCGGCGGC	954
Qy	1426	TGGTACACCTCTCTGAGCACCCCGGTGGCGAGCGCTGACGAATCAGTGCACGCCACCAACC	1485

DD 555 TGGTACACCTTGGCTCAGCAGCCCGTTCGGCGCTTTCGAACTATGCCCACCTCCAGCACC TGTAT

QY 1480 ACACATCCCGGACCGGATCGACGGGACCGCGCTCCATGGATGGACGCGCGGATAC 1544
Db 1015 ACCCTTCGGACCGCACAGTCAACGGCAGTTTCGTACTCCGTGGACGGCGGGCGCTAT 1074

Qy	1740	GAGATCAGGCTCGCATATTTTCATGGGACACGGCAGGAACTGGGAGTCTCGGTGGGCGC	11340	GAGATCAGGCTCGCATATTTTCATGGGACACGGCAGGAACTGGGAGTCTCGGTGGGCGC	11340
Db	1075	GAACTGGAACTCGATATTTTCATGGGACACGGCAGGAACTGGGAGTCTCGGTGGGCGC	11340	GAGATCAGGCTCGCATATTTTCATGGGACACGGCAGGAACTGGGAGTCTCGGTGGGCGC	11340
Qy	1606	TCCCCGACGAGAAACCCGGGCACACGAACATCGGCAAGTACGGAGACGACCTCTAGCTGCAC	1665	TCCCCGACGAGAAACCCGGGCACACGAACATCGGCAAGTACGGAGACGACCTCTAGCTGCAC	1665
Db	1135	TCGTCCGATGGCAGCCCATACGAACTCGGCAAAATACGGGTGACGAGTTGTACSTCGAT	1194	TCGTCCGATGGCAGCCCATACGAACTCGGCAAAATACGGGTGACGAGTTGTACSTCGAT	1194
Qy	1666	CGAGGACCTCCGACACTCCGCGGGTACTCGGTTCGGCCCTACTTCGCGAGCCGCGCCGCC	1725	CGAGGACCTCCGACACTCCGCGGGTACTCGGTTCGGCCCTACTTCGCGAGCCGCGCCGCC	1725
Db	1195	CGCGCATCTTCGGAGCAAGCGTTTATGCGCTGGCACCTCTACACCGCGCGCGCGCGCC	1254	CGCGCATCTTCGGAGCAAGCGTTTATGCGCTGGCACCTCTACACCGCGCGCGCGCGCC	1254
Qy	1726	ATCGACCCCGCGCCCGGATCCGTGACACTGCGCATCTCTCGTCGACACCCAGACGCTCGAG	1785	ATCGACCCCGCGCCCGGATCCGTGACACTGCGCATCTCTCGTCGACACCCAGACGCTCGAG	1785
Db	1255	ATCGATCGGAAACCGCAGATCCGTCCACCTGCGCATCTTTGTAGACACCCAAAGTTGTAG	1314	ATCGATCGGAAACCGCAGATCCGTCCACCTGCGCATCTTTGTAGACACCCAAAGTTGTAG	1314
Qy	1786	GTCTTCGTCAACCGCGGGCCACACCGTGTCTCTCCAGCAGGTCCACTTCGCGCAGGCGCAC	1845	GTCTTCGTCAACCGCGGGCCACACCGTGTCTCTCCAGCAGGTCCACTTCGCGCAGGCGCAC	1845
Db	1315	GTGTTCTGTAATTTCCGGGCACACGCTGGTTTTCGACAGCAGTGCATTCGCGCGCGGGAC	1374	GTGTTCTGTAATTTCCGGGCACACGCTGGTTTTCGACAGCAGTGCATTCGCGCGCGGGAC	1374
Qy	1846	ACGGGAATCTCGCTCTACACCGACGCGCGCCCGCACACTTCACCGGCACTCGCTGCCG	1905	ACGGGAATCTCGCTCTACACCGACGCGCGCCCGCACACTTCACCGGCACTCGCTGCCG	1905
Db	1375	ACGGGATCTCCCTCTATCGGACGCGGTCCGGCCAACTTCACCGGCACTCACCATCCG	1434	ACGGGATCTCCCTCTATCGGACGCGGTCCGGCCAACTTCACCGGCACTCACCATCCG	1434
Qy	1906	GAGATTGGCCAGCGCATC 1923			
Db	1435	GAGTTGGGAACCCCATC 1452			
RESULT 7					
AA	AD17184	AD17184 standard; DNA; 65140 BP.			
XX	AC	AD17184;			
XX	AD	29-NOV-2001 (first entry)			
XX	DE	Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.			
XX	KW	Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;			
KW	KW	antifungal; antibiotic; nys1; ds.			
XX	OS	Streptomyces noursei.			
XX	OS	Location/Qualifiers			
FT	FT	complement (1..1035)			
FT	FT	/*tag= a			
FT	FT	/product= "NysD2 partial protein"			
FT	FT	/note= "CDS does not include stop codon"			
FT	FT	complement (1056..2576)			
FT	FT	/*tag= b			
FT	FT	/product= "NysD1 protein"			
FT	FT	2806..6906			
FT	FT	/*tag= c			
FT	FT	/product= "NysA protein"			
FT	FT	6952..16530			
FT	FT	/*tag= d			
FT	FT	/product= "NysB protein"			
FT	FT	16550..49840			
FT	FT	/*tag= e			

RESULT 7	
AAAD17184	
ID	AAAD17184 standard; DNA; 65140 BP.
XX	XX
XX	AAAD17184;
XX	XX
XX	XX
DT	29-NOV-2001 (first entry)
XX	XX
DE	Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX	XX
XX	Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW	antifungal; antibiotic; nys1; ds.
KW	XX
XX	XX
OS	Streptomyces noursei.
XX	XX

xx	key	Location/Qualifiers
FT	FH	complement (1..1035)
FT	CDS	/*tag= a
FT	FT	/product= "NysD2 partial protein"
FT	FT	/note= "CDS does not include stop codon"
FT	FT	complement (1056..2576)
FT	CDS	/*tag= b
FT	FT	/product= "NysD1 protein"
FT	FT	2806..6906
FT	FT	/*tag= c
FT	FT	/product= "NysA protein"
FT	FT	6952..16530
FT	CDS	/*tag= d
FT	FT	/product= "NysB protein"
FT	FT	16550..49840
FT	FT	/*tag= e
FT	FT	/product= "NysC protein"
FT	FT	

```
FT CDS 50260..51015
FT /tag= f
FT /product= "NysE protein"
FT 51405..54305
FT /tag= g
FT /product= "NysR1 protein"
FT 54329..57190
FT /tag= h
FT /product= "NysR2 protein"
FT /note= "CDS does not include start codon"
FT 57180..59963
FT /tag= i
FT /product= "NysR3 protein"
FT 60415..61047
FT /tag= j
FT /product= "NysR4 (short) protein"
FT /note= "CDS does not include start codon"
FT 61736..62497
FT /tag= k
FT /product= "NysR5 protein"
FT /note= "CDS does not include start codon"
FT complement (62551..63615)
FT /tag= l
FT /product= "ORF2 protein"
FT /note= "CDS does not include start codon"
FT 63765..64961
FT /tag= m
FT /product= "ORF1 protein"
FT
XX WO200159126-A2.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-GB000509.
XX
XX 08-FEB-2000; 2000GB-0002840.
XX 10-APR-2000; 2000GB-0008786.
XX 14-APR-2000; 2000GB-0009387.
XX
XX (UYN0-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
XX (ALPH-) ALPHARMA AS.
XX (SINV-) SINVENT AS.
XX (DZIE/) DZIEGLEWSKA H.
XX (ZOTC/) ZOTCHEV S B.
XX (SEKU/) SEKUROVA O N.
XX (FJAE/) FJAERVIK E.
XX (BRAU/) BRAUTASET T.
XX (STRO/) STROM A R.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
XX WPI: 2001-557614/62.
XX P-PSDB: AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
XX AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
XX useful as antibiotics and antifungals -
XX
XX Claim 2; Page 116-151; 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrolide antibiotic nystatin.
XX The nystatin PKS is useful as antifungal antibiotics. The present
XX sequence is a Streptomyces noursei nysI DNA of nystatin PKS gene cluster.
XX
XX Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;
XX
XX Query Match 7.4%; Score 147.6; DB 22; Length 65140;
XX Best Local Similarity 45.4%; Pred. No. 2.3e-15;
XX Matches 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;
```


Db 80064 ACAGCGGACACACCGCCGCGCTTCGACCTACACCGCGCCCTGCTGGACGCGCAC 80123
Qy 68 TCCTCAGCCTAGACGGGCCCCCTCTCTCGAGGCTCTTCGTCGGGACGCTGAGGCGACTGCGT 127
Db 80124 AACAGCGCGCGCTACCGCGACCTTCGGCGCATCAGCGCGGCGGCGCTGCGGTTCGGCT 80183
Qy 128 CGAAGCTGCTCTCTCTGGGGCGCGCGGTGTGACCGCGAGCTTCGAGACGACAGCGCAG 187
Db 80184 GGAAGGCGTCTCGCTCGCGCGCGCGCGCACACCGTTCGCGCGCGGATCGCGCCCG 80243
Qy 188 GAACGCTCAGTACCGGATCGAGCTCGAGGCGCGCGAGGCTTCAGCGCCCTGAC 247
Db 80244 CCGGAGGACACCGTCAACATCGCGCTTACAGCGCGCGCGGCGGACCGTGTGTCG 80303
Qy 248 CTGCGCGCTTCGCGCTTCAGAGCGCTCCACCGCGGACAGCTCTCTCTTCTACCGCTGCC 307
Db 80304 TCGACTCCCTGCTCCCGCGAGGTCCTCCCGCGAGCACCGCGCGCGCGCGCGCTCC 80363
Qy 308 CGAACGAGGTGAGGCTTCGTCGGCGCCACCGGTTCACGAGAGAACACGAAATGACGC 367
Db 80364 ACCGCGACTCTCTTCACAGTCGAGTGGACCGCGCTCCAGGCGCGCGCGCGCGCGCAC 80423
Qy 368 CGGCATCTACCGCGCGCGCTTCAGAGGAGCGCGCGCGGAGCACTTCGCGCTGATCT 427
Db 80424 CGGCA-----CCGTGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 487
Qy 428 TCGCGGTGCTTCGCGCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 80525
Db 80475 TCCGCG-----CCACGCGCATCCGAGCACCGCGCGCGCGCGCGCGCGCGCGCTCG 80525
Qy 488 ACCATGACCGCGCGCGCGCTTCAGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 547
Db 80526 CCGAGCGGAGGCG 80582
Qy 548 CGCCTACACGCTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 607
Db 80583 CG 80642
Qy 608 ACAGGACACG 667
Db 80643 AGTGCTCG 80702
Qy 668 CCGACTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 727
Db 80703 CCGAGCGGACCG 80762
Qy 728 CGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 787
Db 80763 AGACCGCGCGGTTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTC 80822
Qy 788 AGCAGTACCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 847
Db 80823 AGACCTGGCGACCG 80879
Qy 848 TCATGCTCAACACGAGGTCG 907
Db 80880 GCACCGAGTCAACCG 80939
Qy 908 GTTTCG 967
Db 80940 GGAACCGGACCGACCGCTCTGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80999
Qy 968 GACGACTCGGTAGCGCGCTTACACCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1027
Db 81000 CCG 81059
Qy 1028 GCACTTGAATACCGGACCG 1087
Db 81060 CGGCG 81119
Qy 1088 TCACGCGAGACGGGACCG 1147

Db 81120 ACATCGCGCGCTGGAGACGCTGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81179
Qy 1148 TCGGCTCCCATGACGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACG 1207
Db 81180 CGGCGGCGACCGCGCTACCGCTACCGCTACCGCTACCGCTACCGCTACCGCTACCGCT 81239
Qy 1208 CCGAGGACTCACCGCGCAATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1267
Db 81240 TCGGCTCCCTACCG 81299
Qy 1268 GGCCATCATGACG 1327
Db 81300 GGCACCTGACGAGCG 81359
Qy 1328 AGTAGCG 1387
Db 81360 TCG 81419
Qy 1388 TCGTCGCGGAGCTGGGCTCG 1445
Db 81420 ACGCGCTCG 81479
Qy 1446 CCGGTGCG 1505
Db 81480 CGTGACCG 81537
Qy 1506 CGAGCG 1565
Db 81538 CGCGCGCTCG 81597
Qy 1566 CTGGGACACCG 1622
Db 81598 CCG 81657
Qy 1623 GCACAGCAACATCG 1682
Db 81658 CGAGCG 81717
Qy 1683 CGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1736
Db 81718 CGCGCGCTTCGCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81777
Qy 1737 CGCGCGGATCG 1796
Db 81778 CG 81837
Qy 1797 CG 1856
Db 81838 CG 81897
Qy 1857 GCTCTACACCG 1904
Db 81898 CACCG 81945

RESULT 9

AA53491/c

ID AA53491 standard; DNA; 114955 BP.

XX AC AA53491;

XX DT 05-JUL-1999 (first entry)

XX DE Human adenosine A1 receptor antisense oligonucleotide fragment.

XX DE Antisense oligonucleotide; multiple target; antisense treatment;

KW impaired respiration; inflammation; lung disease;

KW pulmonary vasoconstriction; inflammation; allergic rhinitis;

KW acute asthma; allergy; asthma; impeded respiration;

KW respiratory distress syndrome; pain; cystic fibrosis;

KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;

KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;

KW colon cancer; breast cancer; lung cancer; pancreatic cancer;

Db 103863 CGGCCGAGCGCCGCAVAGACGNNHNNNSGCCGCGCCGAGCGCCGCAVAGACGCGNNHNN 103804
QY 1454 CAGCGCTGACGAATAGCTACCGCCACACACACTCCCGAGCGGACCGCTCGACGCA 1513
Db 103803 NSGCCGCGCGAGCGCCGCAVAGACGCGNNHNNNSGCCGCGCGGCGGCGCAVAGCG 103744
QY 1514 GCGCGCTGCTGCTATGAGGAGCGCGCATACGAGATCGAGTCTGAGTCGATCGGAGCA 1573
Db 103743 GCCGNNHNNNSGCCGCGCGCGCAVAGACGCGCGNNHNNNSGCCGCGCGCGGCGG 103684
QY 1574 CCGCGAGCAAGCTGCGCATCTCGTGTGGCGGCTCCCGGAGCGGAACCGCGCACAGACA 1633
Db 103683 CCGCAVAGCGCGCGCGNNHNNNSACGCGCGCGNNHNNNSGAGCGCGCGCGNNHNNNSV 103624
QY 1634 TCGCAAGTACGAGACAGCTGTACCTCGACCGAGGACCTCCGACCTCGCGGGTACT 1693
Db 103623 GACGCGCGCGNNHNNNSAVGACGCGCGCGNNHNNNSCAVAGCGCGCGCGNNHNNNSC 103564
QY 1694 CGCTGCGCGCGCTACTCGGAGCGCGCGCGCGCATCGACCGCGCGCGCGATCGGTGCA 1753
Db 103563 CAVGACGCGCGCGCGNNHNNNSGCGCAVAGACGCGCGCGNNHNNNSCGCAVAGACGCGCGG 103504
QY 1754 TGCATCTCTGTCGACACCGAGCGCTGAGGTCTTCGTCAACGCGCGCGCACACCGTGC 1813
Db 103503 CNNHNNNSCGCGCAVAGCGCGCGCGNNHNNNSGCGCGCAVAGCGCGCGCGNNHNNNSC 103444
QY 1814 TCTCCAGAGGTCCACTTCGCGGAGCGCGACACGCGGAATCTCGCTCTACACCGACGCG 1873
Db 103443 GCGCAVAGCGCGCGCGNNHNNNSACGCGCGCAVAGACGCGCGCGNNHNNNSGACGC 103384
QY 1874 GCGCGGACACTTACCGCGCATCTGCTCGCGGAGATGCGCGGCGATCTAGCGGATGC 1933
Db 103383 GCGCAVAGCGCGCGCGNNHNNNSGCGCGCGCAVAGACGCGCGCGCGNNHNNNSCGGCG 103324
QY 1934 ACACACACCGCTCACCGAGCGCGCGCG 1964
Db 103323 CGCGCAVAGCGCGCGCGNNHNNNSCGCGAGC 103293

RESULT 10

AAF25795

ID AAF25795 standard; DNA; 3849 BP.

XX AAF25795;

DT 09-APR-2001 (first entry)

XX S. chrysomallus acm-C DNA.

DE Peptide synthetase; PPS; actinomycin synthetase; acm-C; penicillin;

KW N-methyltransferase domain; activation domain; amino acid synthesis;

KW vancomycin; cephalosporin; pristinamycin; actinomycin D; ds.

XX Streptomyces chrysomallus.

XX DE19928313-Al.

XX 21-DEC-2000.

XX 16-JUN-1999; 99DE-1028313.

XX 16-JUN-1999; 99DE-1028313.

XX (KELL/) KELLER U.

XX Schauwecker F, Keller U;

XX WPI; 2001-081744/10.

XX Introducing N-methyltransferase activity into peptide synthetase

XX enzymes, useful for synthesis of N-methylated peptides such as

XX antibiotics, by altering DNA

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Example 2; Page 9; 18pp; German.

This invention describes a novel method of introducing an N-methyltransferase domain (A) into peptide synthetase (PPS) activation domains by altering the DNA that encodes the domains. The invention also describes a method for combining genes (or gene segments) that encode PPS modules that lack (A) with genes (or segments) encoding modules that include (A). Modified DNA sequences formed by insertion of (A) are used (i) for altering natural (or already altered) PPS or polyketide synthetase (PKS) genes and their fragments, also for construction of new PPS and PKS gene and (ii) for construction of plasmids or genetically altered organisms for synthesis of encoded proteins (B). (B) are used for in vivo or in vitro enzymatic synthesis of amino acids, polypeptides, and peptidyl-acyl mixed structures containing N-methylated amino acids, or their derivatives, also for fermentative production of such compounds. These compounds are often of pharmaceutical value, e.g. penicillin, vancomycin, cephalosporin, pristinamycin or actinomycin D. The modified PPS (enzymes) involved in non-ribosomal peptide synthesis) are able to N-methylate their substrates but substrate specificity remains unchanged.

Sequence 3849 BP; 596 A; 1685 C; 1122 G; 446 T; 0 other;

Query Match 6.0%; Score 119.8; DB 22; Length 3849;

Best Local Similarity 44.6%; Pred. No. 7.9e-11;

Matches 558; Conservative 0; Mismatches 687; Indels 6; Gaps 2;

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

Qy	715	ACGCAGAGGTTTCGGCGCCGGCGCGTCTGCGGCTTCGGACACCGACCGACCGCGGT- 773
Db	9005	TGGACCGGCTGCTTTCGGGACACCGCGCCGGCCGCGACGACCTGCCACCTTACGCTTC 9064
Qy	774	-----CCGCAAGATTACACGAGCAGTACCTCTACTTGGTGCACGACGCGGGTTTACAGTTC 828
Db	9065	9124
Qy	829	ACCGCCCTTGCOCGACCCCGTTCATCGTCAACACACCGAGCTGCGCGCCACACACGCCCGCC 888
Db	9125	9184
Qy	889	GAGATCGAGAACGGCGAGTGTTCGCGACCCCAAGATCCATCTAGTGGGACACCGCCCGCGGA 948
Db	9185	9239
Qy	949	GAATGGGTCTCGGTTCATCGGACGACTCGGTACGCCGGTTCACACCTCGCCGAACTCG 1008
Db	9240	9299
Qy	1009	CGCGACTGGACACTTCGCGCAACTTCGACTACCGGAACACCGCCCTCGGGGCACTCGAG 1068
Db	9299	9358
Qy	1069	TGCCCCGACCTGTTCGAGATCACCGCAGACGACGGGACACGCCACTGGGTGCTCGCGGCC 1128
Db	9359	9418
Qy	1129	AGCATGGAGCCTACGGGATCGGCTCCCCATGACGTACGGCTACTGACAGGACACTGG 1188
Db	9419	9478
Qy	1189	GACGCGAGCAGTTTCCACGCGCAGCACCCTACCCCGCAATGGCTCGACTGGGGCTGGGAC 1248
Db	9479	9536
Qy	1249	TGGTACGCGCGCGTCACTGTGCCATCGATGACGCGCCCGAGACAAAGCGCTCGCCATC 1308
Db	9537	9596
Qy	1309	CGTGTGATGACAACTGGAAGTACGCCACGCGACGTCCCCACGGAGCGCATCGACGGC 1368
Db	9597	CCCCGGCGCGCGCGCAACCGCTGCGCTTGGACGGCCACTACGACGCGCTC-CCGACGCGC 9655
Qy	1369	TACAACGGCAGAACTCGATCGTCCGGAGCTGCGGCTGCGCCGACACGCTTGGGGCTGG 1428
Db	9656	9715
Qy	1429	TACACCTCTCTAGACACCCCGTGGCAGCGTGCAGAACTACGTACCGCCACCACCAACA 1488
Db	9716	GTCTACGCGGAGTCTGAGTCCCGAGCGCGCGGTCGCGACGCGGAGGCGTTTCGGCCTC 9775
Qy	1489	CTCCCGACGCGCGCTGACGGCAGCGCCCTCTCGCATGGACGGACGCGCATACGAG 1548
Db	9776	9835
Qy	1549	ATCAGAGTCTGACATCGCTTGGGACACCGCAGAACTCGCGCATCTCGTGGCGCGCTCC 1608
Db	9836	GGCCGGGCGCCTTCCGCTTCTCTGGAGGGCGTCTCCCTCCAGCGCGCGCGCGCCACC 9895
Qy	1609	CCGACAGGAACCCGGCACACGAACATCGGCAAGTACGGAGACAGCCTGTACGTGCAACGA 1668
Db	9896	9955
Qy	1669	GGACCCCTCGACCTCGCGGGTACTCGCTCGCCCCCTACTCGGAGCGCGCGCCCCCATC 1728
Db	9956	10015
Qy	1729	GACCCGCGC---GCCGATCGGTCACTCGGCATCTCTCGTCGACACCCAGAGCGTCGAG 1785
Db	10016	10075
Qy	1786	GTCTTCTCAACGCGCGGCACACCGTGTCTCTCCAGCAGAGGTTCACCTTTCGCGCGGGCGAC 1845

[illegible]

RESULT 13
AAI99683/0

AAI99683/C
ID AAI99683 standard: DNA: 4403765 BP.

AA
AC
AAI99683:

15-JAN-2002 (first entry)

Mycobacterium tuberculosis

XX
KW Mycobacterium tuberculosis. strain H37Pv. strain CDC 1551.

variation; epidemiology; patient treatment; epidemic monitoring; ds.
KW
YY

OS Mycobacterium tuberculosis.

PN US6294328-B1.

PD 25-SEP-2001.

PF 24-JUN-1998;

PR 24 - JUN - 1998; 98US - 0103840.

PA (GENO-) INST GENOMIC RES.

PI Fleischmann RD, White OF

DR WPI: 2001-647261/74.

PT Evaluating strain va

determining the nucleotide sequence of the strain at PT genome corresponding to positions where *M. tuberculosis* PT

PT 1551 and H3/RV differ - XX

PS Claim 4; SEQ ID NO 2; 3pp + sequence listing; English.
XX

CC The invention relates to evaluating strain variation within
CC different populations of the tubercle bacillus, *Mycobacterium tuberculosis*, and

CC *Mycobacterium tuberculosis* or related *Mycobacterium tuberculosis* complex
CC nucleotide sequence of the first strain at positions 1-1000

CC sequence of the genome that correspond to positions that differ in the

CC H37Rv (AAI99682). The method is useful for evaluating strain variation of

CC tuberculosis genetics, epidemiology, patient treatment and epidemic

CC Note: The sequence data for this patent did not form part of the printed

CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 5.0%; Score 101; DB 22; Length 4403765;

BEST LOCAL SIMILARITY 42.3%, FILED: NO. 4.1E-08,
 Matches 833: Conservative 0: Mismatches 1115: Indels 23: Gaps 4

0v 11 CCGACTTCCCCCTCGACGACCAACGGTCCCGCTACCGGGCGACCGCCGGGCGGACGTGCTCC 70

Db 3941672 CCGCGGGCCCCCGGACACCGGGCGTGGCTGCTGGCGGATCATGGCGCGCGGTGGCTGG 3941611

0.7 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130

Query Match 5.0%; Score 101; DB 22; Length 4403765;
Best Local Similarity 42.3%; Pred. No. 4.1e-08;
Matches 833; Conservative 0; Mismatches 115; Indels 23; Gaps 4

Qy 11 CCGACTTCCCTCGACGACACCGTCCCCCTACGGGGCGACCGCCCGGCGACTGCTCC 70

Db 3941672 CCGCGGGCCCCGCGACACCGGCGTTGCCTGTGGTGCGGATCATGCCGCGGGTCCCTCG 394161

71 TCAGCCTAGACGGGGCCCCCTCCTCGAGGCTTTCGTCGGGGACGGTGAGGGACTGCGTCA 130

[illegible]

RESULT 14	
AAAX53491	
ID	AAAX53491 standard; DNA; 114955 BP.
XX	
XX	
AC	
XX	AAAX53491;
XX	
DT	05-JUL-1999 (first entry)
XX	
XX	Human adenosine A1 receptor antisense oligonucleotide fragment.
DE	
XX	
XX	Antisense oligonucleotide; multiple target; antisense treatment;
KW	impaired respiration; inflammation; lung disease;
KW	pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW	acute asthma; allergy; asthma; impeded respiration;
KW	respiratory distress syndrome; pain; cystic fibrosis;
KW	pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW	chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW	colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW	

KW	hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW	prostate cancer; ss.
XX	Synthetic.
OS	
XX	
PN	WO9913866-A1.
XX	
PD	25-MAR-1999.
XX	
PB	17-SEP-1998; 98WO-US19419.
PF	
XX	
PR	09-JUN-1998; 98US-0093972.
PR	17-SEP-1997; 97US-0059160.
XX	
PA	(UYEC-) UNIV EAST CAROLINA.
XX	
PI	Nyce JW;
XX	
DR	WPT; 1999-229400/19.
XX	
PT	New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT	vasoconstriction
PS	Disclosure; Page 37; 120pp; English.
XX	
CC	The specification describes antisense oligonucleotides (AA52869-X55271)
CC	directed against at least 2 mRNAs selected from target genes, coding and
CC	non-coding regions of RNAs corresponding to target genes, gene
CC	initiation codons, genomic flanking regions, intron-exon borders, the
CC	5'-end, the 3'-end and the juxta-section between coding and non-coding
CC	regions and all segments of RNAs encoding proteins associated with one
CC	or more diseases, conditions or mixtures. The antisense oligonucleotides
CC	may be derived from sequences AA55272-74. These multiple target
CC	oligonucleotides (specifically AA55180-271) can be used for the
CC	antisense treatment of diseases and conditions. Typical diseases and
CC	conditions are those associated with impaired respiration and
CC	inflammation, including lung diseases, pulmonary vasoconstriction,
CC	respiratory distress syndrome, pain, cystic fibrosis,
CC	asthma, allergies, acute asthma, chronic bronchitis, emphysema, chronic
CC	pulmonary hypertension, obstructive pulmonary disease (COPD), and cancers such as leukemias,
CC	lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,
CC	pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,
CC	hepatic metastases, as well as all types of cancers which may metastasize
CC	or have metastasized to the lungs, including breast and prostate cancer.
XX	
SQ	Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 Other;
	Query Match 5.0%; Score 99.8; DB 20; Length 114955;
	Best Local Similarity 34.0%; Pred. No. 8.9e-08;
	Matches 604; Conservative 118; Mismatches 1032; Indels 21; Gaps 7;
QY	2 CGGTGACCGCCAGTTCCTTCGAGCACCGTGCTCCCCCTACCGGCCGACCGCCGGCCC 61
	: : : : : : : : : : :
Db	104356 CBGGCGCGCGCGCGCGGGSNNNDNCCBGCBGGCGCGCGCGCGCGCGGCSNNN 104415
	: : : : : : : : : : :
QY	62 GACTGCTCCTCAGCCTAGACGGGCGCCCTCTCTGAGGTCTTCTGGGGACGCTGAGCGGA 121
	: : : : : : : : : : :
Db	104416 DNNCCGBGGCCBGGCGCGCGCGGCGGGSNNNDNCCBGCBGGCGCGCGCGCGCGG 104475
	: : : : : : : : : : :
QY	122 CTGCCTGAACCTTGTCTCTTGGGGGCGCGGTGTGACCGCGAGCTCTGAGACGGCAC 181
	: : : : : : : : : : :
Db	104476 GCCSNNNDNCCBGCGCCBGGCGCGCGCGGCGGCSNNNDNCCBGCBGGCGCGCGCGG 104535
	: : : : : : : : : : :
QY	182 GGCCAGGAACCGTAGCAGTCAGCGGATCTGAGTTCAGGCGCGCGCGCGCGCGCGCGCG 241
	: : : : : : : : : : :
Db	104536 CGCGGGSNNNDNCCBGCGCCBGGCGCGCGCGCGGCGGCSNNNDNCCBGCBGGCGCGCGG 104595
	: : : : : : : : : : :
QY	242 CTGAACCTGCGCGCTTCTGCGCTGACGAGCGCTCCACCGCGCAGCTCTCTTCTACC 301
	: : : : : : : : : : :
Db	104596 CGCGSNNNDNCCBGCGCCBGGCGCGCGCGGCGGCSNNNDNCCBGCBGGCGCGCGCGG 104655
	: : : : : : : : : : :
QY	302 GCTCCCCGAACAGGTTGAGCGTCTGTCTGGCGCCACCGCTCCAGAGAACCAAGCAA 361
	: : : : : : : : : : :

```
Db 105728 GGGCCCCCTGGCTGGCGCCCGCGGCTTGGCCCGCCCGGCGSNNNDNNGCGCCCTGGC 105787
QY 1432 ACCCTCTGAGCACCCCTGGCGAGCGCTGACGAACTAGCTACCGCCACACACACTC 1491
Db 105788 TCGGCCCCCGCGCGCTTGGCCCGCGSNNNDNNGCGCCCTGGCTGGCGCCCGCG 105847
QY 1492 CCGGACCGGACGTCGACGCGGAGCGCGCTGCTGGCCATGG---AACGGACGCGCATACGAG 1548
Db 105848 CCGCGCTTCCGCGCGCGSNNNDNNGCGCCCTGGCTGGCGCCCGCGGCTTGGCCG 105907
QY 1549 ATCGAGCTCGACATCGCTTGGGACACCGGACGAGCTGCGATCTCGTGGCGCGCTCC 1608
Db 105908 CCSSNNNDNNGCGCCCTGGCTGGCGCCCGCGGCTTGGCCCGCCCGSNNNDNNGGCGC 105967
QY 1609 CCGGACGGAACCGGCACACGACATCGCAAGTAGCGGAGCAGACCTGTACGTGACCGCA 1668
Db 105968 CTGGGCTCGGCGCGCGCGCTTGGCCCGCGSNNNDNNGCGCCCTGGCTGGCGCCCGC 106027
QY 1669 GAGCCTCGGACCTCGCCGGGTACTGCTGGCCCGCTACTCGGAGCGCCCGCCCGCATC 1728
Db 106028 GCGCGGCTTGGCGCGSNNNDNNGCGCCCTGGCTGGCGCCCGCGGCTTGGCCSNN 106087
QY 1729 GACCCCGCGCGCGATCGCTGACCTGCGCATCCT 1763
Db 106088 NDNNGGCCCCCTGGCTGGCGCCCGCGGCGCGGCTT 106122

RESULT 15
AA14651
ID AA14651 standard; DNA; 77536 BP.
XX
AC AA14651;
XX
DT 08-AUG-2000 (first entry)
XX
DE Nucleotide sequence of the FK-520 biosynthetic gene cluster.
XX
KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW Streptomyces hygroscopicus var. ascomyceticus; immunophilin;
KW FK-506 binding protein; polyketide compound; transplant rejection;
KW graft-versus-host disease; uveitis; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.
XX
OS Streptomyces hygroscopicus.
XX
FH
FH Key Location/Qualifiers
FH CDS complement (412..1836)
FT /*tag= a
FT /note= "fkwb gene"
FT complement (2020...3579)
FT /*tag= b
FT /note= "fkvb gene"
FT 3969..4496
FT /*tag= c
FT /note= "fkbr2 gene"
FT complement (4595..5488)
FT /*tag= d
FT /note= "fkbr1 gene"
FT 5601..6818
FT /*tag= e
FT /note= "fkbe gene"
FT 6808..8052
FT /*tag= f
FT /note= "fkbf gene"
FT 8156..8824
FT /*tag= g
FT /note= "fkbg gene"
FT complement (9122..9883)
FT /*tag= h
FT
FT CDS
FT /note= "fkBH gene"
FT complement (9894..10994)
FT /*tag= i
FT /note= "fkbi gene"
FT complement (10987..11247)
FT /*tag= j
FT /note= "fkBj gene"
FT complement (11244..12092)
FT /*tag= k
FT /note= "fkBk gene"
FT complement (12113..13150)
FT /*tag= l
FT /note= "fkBl gene"
FT complement (13212..23988)
FT /*tag= m
FT /note= "fkBc gene"
FT complement (13452..13662)
FT /*tag= n
FT /note= "ACP6"
FT complement (13761..14394)
FT /*tag= o
FT /note= "KR6"
FT complement (14517..15294)
FT /*tag= p
FT /note= "ER6"
FT complement (15438..16587)
FT /*tag= q
FT /note= "dehydratase domain (DH) 6"
FT complement (16587..17820)
FT /*tag= r
FT /note= "acyltransferase domain (AT) 6"
FT complement (17820..19053)
FT /*tag= s
FT /note= "KS6"
FT complement (19116..19326)
FT /*tag= t
FT /note= "ACP5"
FT complement (19464..20097)
FT /*tag= u
FT /note= "KR5"
FT complement (20241..21420)
FT /*tag= v
FT /note= "DH5"
FT complement (21420..22653)
FT /*tag= w
FT /note= "AT5"
FT complement (22653..23892)
FT /*tag= x
FT /note= "KS5"
FT complement (23992..46573)
FT /*tag= y
FT /note= "fkBB gene"
FT complement (24163..24373)
FT /*tag= z
FT /note= "ACP4"
FT complement (24997..26146)
FT /*tag= aa
FT /note= "DH4 (inactive)"
FT complement (26146..27430)
FT /*tag= ab
FT /note= "AT4"
FT complement (27430..28684)
FT /*tag= ac
FT /note= "KS4"
FT complement (28750..28960)
FT /*tag= ad
FT /note= "ACP3"
FT complement (29092..29740)
FT /*tag= ae
FT /note= "KR3"
FT complement (29869..31018)
FT /*tag= af
FT /note= "DH3 (inactive)"
FT
```

FT	misc_feature	complement (31018..32185)
FT		/tag= ag
FT		/note= "AT3"
FT	misc_feature	complement (32185..33439)
FT		/tag= ah
FT		/note= "KS3"
FT	misc_feature	complement (33505..33715)
FT		/tag= ai
FT		/note= "ACP2"
FT	misc_feature	complement (33823..34480)
FT		/tag= aj
FT		/note= "KR2"
FT	misc_feature	complement (34606..35749)
FT		/tag= ak
FT	misc_feature	complement (28750..28960)
FT		/tag= al
FT		/note= "DH2 (inactive)"
FT	misc_feature	complement (35749..37144)
FT		/tag= am
FT		/note= "AT2"
FT	misc_feature	complement (37145..38296)
FT		/tag= an
FT		/note= "KS2"
FT	misc_feature	complement (38371..38581)
FT		/tag= ao
FT		/note= "ACP1"
FT	misc_feature	complement (38677..39307)
FT		/tag= ap
FT		/note= "KR1"
FT	misc_feature	complement (39442..40609)
FT		/tag= aq
FT		/note= "DH1"
FT	misc_feature	complement (40609..41842)
FT		/tag= ar
FT		/note= "AT1"
FT	misc_feature	complement (41842..43093)
FT		/tag= as
FT		/note= "KS of extender module 1 (KS1)"
FT	misc_feature	complement (43144..43660)
FT		/tag= at
FT		/note= "ACP of loading domain"
FT	misc_feature	complement (43777..44629)
FT		/tag= au
FT		/note= "ER of loading domain"
FT	misc_feature	complement (44974..46573)
FT		/tag= av
FT		/note= "CoA ligase of loading domain"
FT	CDS	46754..47788
FT		/tag= aw
FT		/note= "fkxB gene"
FT	CDS	47785..52272
FT		/tag= ax
FT		/note= "fkbp gene"
FT	CDS	52275..71465
FT		/tag= ay
FT		/note= "fkBA gene"
FT	misc_feature	52362..53576
FT		/tag= az
FT		/note= "KS7"
FT	misc_feature	53577..54716
FT		/tag= bb
FT		/note= "AT7"
FT	misc_feature	54717..55871
FT		/tag= ba
FT		/note= "DH7"
FT	misc_feature	56019..56819
FT		/tag= bc
FT		/note= "ER7"
FT	misc_feature	56943..57575
FT		/tag= bd
FT		/note= "KR7"
FT	misc_feature	57710..57920
FT		/tag= be

[illegible]

Db 62147 GCCCGTCGGCTCCGGGAGGAGTTCGTATGCGCCCGGGATGGACCCCGGCACAGCCGAT 62206
 QY 1371 CAACGGGAGAACTCG---ATCGTCCGGAGCTGGGGTTCGCCGACAGCCCTGGCGGGTG 1427
 Db 62207 GACCGGCTCCGTACCGCCCATCTGAGCGGCTGCGCAGGAGCGCGCGGCTCGCCCG 62266
 QY 1428 GTACACCTCTCTGAGACCCCGTGGCAGCGGTGACGAACCTAGCTACCGGCGACCAACAC 1487
 Db 62267 TGCCGGGAGAGCTTCGCCACGCGCTGCGGAGCTGCGCCGACGCGGACCGCGGCGCGC 62326
 QY 1488 ACTCCCGGACCGACCGTTCGCGGAGCGCGCTCTGTCATGGAACGAGCGGCATACGA 1547
 Db 62327 GGTGACACCTCTGCTCGGAGCCGCGCCCGCTGCTCGGCCACGCGGAGCCCTCCGA 62386
 QY 1548 GATCGAGCTCGACATCGCTGGGACACCGCGACGAAACGTCGGCATCTCGGTGGCGCGTC 1607
 Db 62387 GATCGCGCGACCGAGCTTCAAGGACCTCGGCATCGACTCGCTCACCGGATCGAGCT 62446
 QY 1608 CCCCAGCGAACCGGACACAGATCGGCAAGTACGAGAGACC---TGTAGTCTGA 1664
 Db 62447 GCGCAACCGGCTCGGAGGAGCGGACCGGGCTGCGGCTGAGTGCACGCTGGTCTCGACCA 62506
 QY 1665 CGGAGGACCTCCGACCTCGCGGGTACTCGCTGCGCCCTACTCGCGAGCGCGCGCCC 1724
 Db 62507 CCGACACTCGGGTCTCGCGCCCAAGCTCCGACCCGATCTGTTCCGACGCGCGTGCC 62566
 QY 1725 CATGACCCCGGCGCCGATCCGTGACCTGCGCATCTCGTTCGACACCCAGAGCGTCTGA 1784
 Db 62567 CACGCGCGCGGAGCGCACGACCCACGACGAGCCACTCGCGATCTGTCGGCATGSC 62626
 QY 1785 GGTCTTCGTCAACGCGCGGCGCACCGTCTCTCCAGAGGTTCACCTTCGCGGAGGGCGA 1844
 Db 62627 GTGCGGACTGCGCGCGGGTCCGCTCGCGGAGGACCTGTGGCAGCTCGTGGCGTCCGG 62686
 QY 1845 CACGG 1849
 Db 62687 CACCG 62691

Search completed: May 10, 2003, 07:35:05
 Job time : 4373.5 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 04:44:16 ; Search time 2890 Seconds
(without alignments)
11213.553 Million cell updates/sec

Title: US-09-868-328B-3

Perfect score: 2001

Sequence: 1 ggggtgaccccgactccc.....atcgacacgtctctcgctt 2001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estlin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	94.4	4.7	1288	14	BQ678719
C 2	90.6	4.5	1016	17	CNS03LU9
C 3	88.2	4.4	1046	14	BQ952554
C 4	88	4.4	1651	12	BG809816
5	86.6	4.3	1065	14	BQ681076
6	86.6	4.3	1218	14	BQ876717

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1	BQ678719	1288 bp	mrna	linear	EST 15-JUL-2002
LOCUS	AGENCOURT_8199257	NIH_MGC_112	Homo sapiens	cdna clone	IMAGE:6261563
DEFINITION	5', mRNA sequence.				
ACCESSION	BQ678719				
VERSION	BQ678719.1	GI:21791398			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 1288)				
TITLE	NIH-MGC http://mgs.nci.nih.gov/ .				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabs-r@mail.nih.gov Tissue Procurement: DCTD/DTF cdna Library Preparation: Rubin Laboratory cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCM2422 row: h column: 12 High quality sequence stop: 171. Location/Qualifiers 1. :1288				
FEATURES	source				

7	84.4	4.2	1339	13	BM458211
C 8	84.2	4.2	827	17	AZ183795
9	82.6	4.1	733	17	AG039579
C 10	82	4.1	934	17	AG046262
C 11	82	4.1	1023	17	AG128304
C 12	81.6	4.1	1341	17	AG030611
C 13	81.2	4.1	1242	14	BM911414
14	81.2	4.1	1281	12	BG852363
15	80.8	4.0	1051	14	BQ689718
C 16	80.6	4.0	1116	12	BG810038
C 17	80.6	4.0	1233	17	AG074706
C 18	80.4	4.0	1321	17	AG126084
C 19	79.8	4.0	1798	17	AG171124
C 20	79.4	4.0	1112	14	BQ058195
C 21	78.2	3.9	666	12	BG786337
C 22	78.2	3.9	724	17	AG171201
23	78.2	3.9	1061	17	AG046200
24	78.2	3.9	1150	14	BQ892481
25	78	3.9	775	17	AZ185516
26	78	3.9	821	17	AG063084
C 27	78	3.9	1243	14	BM913931
C 28	77.8	3.9	1143	10	AW731158
C 29	77.6	3.9	715	17	AG036665
C 30	77.6	3.9	1128	14	BQ710735
C 31	77.6	3.9	1170	17	AG111669
C 32	77.2	3.9	1297	14	BQ064843
C 33	77	3.8	1025	17	AG137560
C 34	77	3.8	1123	17	AG080476
C 35	77	3.8	1134	17	AG043044
C 36	77	3.8	1137	12	BG809979
C 37	76.8	3.8	1268	13	BM551006
C 38	76.6	3.8	1127	14	BQ718629
C 39	76.4	3.8	763	17	AG083422
C 40	76.4	3.8	1197	13	BI416470
C 41	76.2	3.8	1138	9	AI374006
C 42	76	3.8	1149	14	BM910848
C 43	76	3.8	1798	17	AG171124
C 44	75.8	3.8	1250	13	BM560215
C 45	75.8	3.8	1424	14	BQ720687


```
Qy 281 CCCGACAGCTCTCTTACCGCTCCCGAACCAGGGTGGAGCTTCTGTCGCCGCCACCC 340
Dy 892 CCCCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 833
Qy 341 GTCCACGAGAGAACCAACATGAGCCCGGCAATCTACGCGCGCGCGCTGCTCCAGGAG 400
Dy 832 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 773
Qy 401 CCGCGCCCGGAGCACTCCCTGATCTTCTGCGGTGCTGTCGCGCTGACGCGCGGCAT 460
Dy 772 CGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 713
Qy 461 CCGTTCGGGTGCTGCTCGTCCGCTTACACATGACGCCGCCCGCCCGCGCTGCTCGG 520
Dy 712 GCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 653
Qy 521 ACCCCACGCGCGCTGACACCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 580
Dy 652 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 593
Qy 581 AGAACACGCGCGCTGGGACCAACGAGACGACGCGCTGCTTACGCG 640
Dy 592 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 533
Qy 641 ACCAGGCAACGATGCGTGGGCGCCGACTTCCCGCTGTGTCGCGGTGCGGTCG 700
Dy 532 GCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 473
Qy 701 TCGCACCGGACACGACGAGGTTCGGCGCGCGGGGCGGTGTCGCGTTCGGACCG 760
Dy 472 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 413
Qy 761 CGACGCGCGCTCGCAAGTACAGGAGCAGTCTTACTGTGTCGACGACGCGGCT 820
Dy 412 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 353
Qy 821 TCAGTTTACCGCTGCGCCGACCGCGTATGTCGTAACACGAGGTCGCGCGCCACCA 880
Dy 352 GCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 293
Qy 881 CGCGCGCGGAGATGAGAACCGCGAGTGTTCGCGACCGCCCAAGATCCACTGGACCG 940
Dy 292 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 233
Qy 941 CCGCGGAGATGCTGCTGCTATCGGACGACTCGGTACGCGCGCTTCTACACTCGC 1000
Dy 232 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 173
Qy 1001 CGACCTGCGGCTGAGACATTCGCGCAACTTCGACTACCGAACACGCGCTCGCG 1060
Dy 172 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 113
Qy 1061 GCATGAGTGCCTGCTGCTGAGATCAGCGAGACGAGGACAGCGCAC 1113
Dy 112 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
```

```
RESULT 3
BO952554/c BO952554 1046 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8863775 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6423858
DEFINITION 5', mRNA sequence.
ACCESSION BO952554
VERSION BO952554.1 GI:22368032
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1046)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
```

COMMENT Contact: Robert Strausberg, Ph.D.
Email: ccapbs-r@mail.nih.gov
Tissue Procurement: DCTD/BMP/Gazdar
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM2603 row: b column: 19
High quality sequence stop: 206.
Location/Qualifiers
1. .1046
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6423858"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

FEATURES
source

BASE COUNT 24 a 324 c 538 g 30 t 30 others
ORIGIN
Query Match 4.4%; Score 88.2; DB 14; Length 1046;
Best Local Similarity 45.6%; Pred. No. 1.7e-05;
Matches 379; Conservative 0; Mismatches 436; Indels 16; Gaps 3;
Qy 1 GCGGTGACCCCGACTTCCCTTCGACGACACCGTCCCTTACGCGCGCGCGCGCCGCC 60
Dy 988 GCG 929
Qy 61 CGACTGCTCTCTAGCTAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Dy 928 GCG 869
Qy 121 ACTGCGTGAACCTGCTCTCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Dy 868 GCG 809
Qy 181 CG----GCCAGGACCGTGCACGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 235
Dy 808 GCG 749
Qy 236 ACGCCCTTGAACCTGCG 295
Dy 748 GCG 689
Qy 296 TCTACCGCTGCGCGAACCGAGGTGGACGCTTCTGTCGCGCGCGCGCGCGCGCGCG 355
Dy 688 CCG 629
Qy 356 CAGCAATGACCG 415
Dy 628 CCG 569
Qy 416 TCGCGCTGATCTTGGCGCGGTGCTGTCGCGCGCTGACGCGCGCGCGCGCGCGCG 468
Dy 568 CCG 509
Qy 469 GGCTGCTTCTGCGGTCTACCATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 528
Dy 508 GCG 449
Qy 529 GCGCGGTACACCGACCGCGCGCGCTTACCGAGTGTACTTACCTGCGACTCGACG 588
Dy 448 CCG 393

```
QY 589 GGGCCCGGGGCTGGGACACAGGACAGACAGGCGGCTGCCCTTACGACACACAGGCG 648
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 392 CGGGCCCGGGCCCGCCCGCCCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 649 ACCGTGATCGGCTGGGGCGGCGGCTTCGCGGCTGGTTCGGGTCGGGTCGTCGCGACC 708
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 332 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 273
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 709 GCGAACACGCGAGGTTGGCGGCGGGCGGGTGTGTCGCGCTCGCGACCCAGCGACCGAC 768
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 272 CCGCGCCCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 213
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 769 GCGCTCGCGAAGTACAGAGGACGAGTACTCTACTGTGTGACCGACCGGCGGG 819
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212 GSGNGNNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 162

RESULT 4
BG809816/c
LOCUS
DEFINITION
  mgct001xk18f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe
  grisea cDNA clone mgct001xk18f 5', mRNA sequence.
ACCESSION
  BG809816
VERSION
  BG809816.1 GI:14180796
KEYWORDS
  EST.
SOURCE
  Magnaporthe grisea.
  Magnaporthe grisea
  Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
  Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
REFERENCE
  1 (bases 1 to 1651)
  Choi, W. and Dean, R.A.
  Construction and sequence analysis of an appressorium stage cDNA
  library in the rice blast fungus, Magnaporthe grisea
  Unpublished (2001)
  Contact: Ralph A. Dean
  Fungal Genomics Laboratory
  North Carolina State University
  Campus Box 7251, Raleigh, NC 27695, USA
  Tel: 919-513-0020
  Fax: 919-513-0024
  Email: ralph.dean@ncsu.edu
  Seq primer: T3 primer (AATTAAACCTCACTAAAGG).
  Location/Qualifiers
    1..1651
      /organism="Magnaporthe grisea"
      /strain="70-15"
      /db_xref="taxon:148305"
      /clone="mgct001xk18f"
      /clone_lib="Magnaporthe grisea Appressorium Stage cDNA"
      /dev_stage="Germinated conidia on appressorium-inductive
      surface"
      /note="Vector: pBlueScript SK(+); Vector: Site_1: EORI;
      Site_2: XhoI; The appressorium formation-specific cDNA
      library was constructed from conidia germinated for 5-8 hr
      on an inductive surface. The library contains over 55,000
      clones with average insert size of 1.5 kbp."
BASE COUNT
  66 a 320 c 769 g 21 t 475 others
ORIGIN
  Query Match 4.4%; Score 88; DB 12; Length 1651;
  Best Local Similarity 31.6%; Pred. No. 1.8e-05;
  Matches 445; Conservative 0; Mismatches 953; Indels 9; Gaps 3;
  QY 602 GGGACACGACGACGACGACGCGGCTGCTTACGACGACGACGACGCGTATGCGCGC 661
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  Db 1570 GCGCGCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 1511
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  QY 662 TCGGCGCGGACGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCG 721
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  Db 1510 CCGCGCGCGCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 1451
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  QY 722 GGTTCGGCGCGGGCGGGTCGTGCGGCTCGCGACCGACCGGCGGCGGCGGCGGCAAGT 781
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


/clone_lib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 63 a 629 c 334 g 95 t 97 others
ORIGIN

Query Match 4.3%; Score 86.6; DB 14; Length 1218;

Best Local Similarity 41.9%; Pred. No. 3 1e-05;

Matches 334; Conservative 0; Mismatches 463; Indels 1; Gaps 1;

QY 7 CACCCGACTTCCCTCGAGACACCGTCCCTTACCGCCGACCGCCGCGCCGACTG 66

DB 419 CCCCCGGGGCGCCNCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 478

QY 67 CTCTCAGCTAGACGGCGCCCTCTCTGAGGTCTTCTGCGGACGCTGAGCGACTGC 126

DB 479 CCCNNGCGNCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 538

QY 127 TCGAACCTGTCTCTCTGGGGCGCGCGGTGTGACCGCGAGCCTCGAGACGGCAGCGCCA 186

DB 539 CCCCCCGCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 598

QY 187 GGAACGCTGACGTGACGGGATGAGCTGAGGGCCCGCAGCGATCTGACGCCCTGAA 246

DB 599 CCGCGCGCGCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 658

QY 247 CTGCGCGCGCTTCTGGCTGACGAGCGCTCCACACCGACAGCTCTCTTACGCGTCG 306

DB 659 CCGCCCNCCCGGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCG 718

QY 307 CCGAACAGGGTGGAGCGCTTCTGTCGCGCCACCGCTCCACGAGAGGAACAGCAATGAG 366

DB 719 CGCGCCCGGCGGNGNGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 778

QY 367 CCGGCCATCTACCGCGCGCGCTGCTCAGAGGAGCGCGCGCGGAGCACTGCGCTGATC 426

DB 779 CCCCCCGCCCGGNGGNCNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 838

QY 427 TCGCGCGGTGCTGCGCGCTGACGCGCGGATCCGCTCCGCGGCTCGCTGCGCGCTC 486

DB 839 CGCNCGCGNCCCG 898

QY 487 TACACATGACGCGCGCGCGCGCTGGCTTGGAGACCCCAAGCCCGGTCACCAACAC 546

DB -899 CCGCCCG 958

QY 547 GCGCGCTACAGCTGACTACCTGCACTCCGACCAAGAAACACGGGCCCGCGCGGCTGGGAC 606

DB 959 GCGCCCG 1018

QY 607 CACGCGAGCAGCAGCAGCGCGCTTTCAGCAGCAGCAGCAGCAGCAGCAGCAGC 666

DB 1019 CGCGCCCG 1077

QY 667 CCGGACTTCCCGGTGTGGTTCGGGGTTCGGGGTTCGGGGTTCGGGGTTCGGGGTTC 726

DB 1078 CGCCCG 1137

QY 727 GCGCGCGCGCGGTCTCGCGCTCGCGACCGACCGCGCGCGCGCGCGCGCGCGCG 786

DB 1138 GCGCCCG 1197

QY 787 GAGCAGTACTTCTACTGG 804

DB 1198 GCG 1215

RESULT 7

BM458211

LOCUS

DEFINITION

AGENCY

5', mRNA sequence.

BM458211

VERSION

BM458211.1. GI:18507251

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1339)

NIH-MGC http://mgi.nhlbi.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM12210 row: o column: 01

High quality sequence stop: 134.

Location/Qualifiers

1..1339

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5530584"

/clone_lib="NIH-MGC_71"

/tissue_type="leiomyosarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: NotI;

Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 2.1 kb.

BASE COUNT 63 a 780 c 403 g 78 t 15 others

ORIGIN

Query Match 4.2%; Score 84.4; DB 13; Length 1339;

Best Local Similarity 47.9%; Pred. No. 7.4e-05;

Matches 414; Conservative 0; Mismatches 426; Indels 24; Gaps 5;

QY 2 CGGTGACCCGACCTTCCCTCGAGACACCGTCCCTACCGCGCGCGCGCGCGCGCG 61

DB 253 CG 312

QY 62 GACTGCTCTCAGCTTAGACGGGCGCCCTCTCTCGAGGTCTTCTCGGGGACGGTGA 121

DB 313 GCG 372

QY 122 CTGCGTGAACCTGTGCTCTGCGGGCGCGGGTGTACCGGAGCTCTGAGAGAGAC 181

DB 373 CG 432

QY 182 GCGCAGGAACCTGACGCTGACCGATGACGTGAGCGCGCGCGCGCGCGCGCGCG 241

DB 433 GCG 492

QY 242 CTGAACCTGCGCGCTTCTGGCTGACGAGCGCTCCACCGCGAGAGTCTCTTTTAC 301

DB 493 CG 544

QY 302 GCTGCCGGAACAGGGTGGACGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 361

DB 545 GCG 604

QY 362 TGACGCGCGCGCATCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 421

Db	820	GNNGNNGGCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCG	761
Qy	571	CACCTCCGACCAGAACAA- - - - CGCGCCCGCCGGCTGGGACCAAGCAGCAGCAGCAGCAGCG	626
Db	760	CGCNGNGCGCGCGCNGNGCGCGCCCGCGCCCGCCCGCCCGCCCGCCCGC- - - - -	701
Qy	627	CGTCGCGCTTCACGACGACACAGGCAACGCTGATGTCGTCGCGCCCGCACTCCCGCTGTGGTTC	686
Db	700	NNCGCCGCNCNCGCCGCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	641
Qy	587	CGGGTCGGCGCGTCTGTCGGCACCGCGACCGCAGGGTTCCGGCGCGCGCGCGCTCGTCGC	746
Db	640	GG-GCG	581
Qy	747	GC-TCGCGACCCAGCGACGACGCGGTCCGCAAGTACCAGGAGCAGTACCTCTACTTGGT	805
Db	580	GCGCGCGCGCCCGCNGCG	521
Qy	806	CGACCGACGCGCGGTTCACGTTACCGCGCCCTGCCCAGCCCGCTATCGTCAACACCGACG	865
Db	520	GCG	461
Qy	866	GTGCGCGCGCCACACGCGCCGCGGAG	891
Db	460	GCGCGCGCGCCCG	435
RESULT 13			
LOCUS	BM911414		
DEFINITION	AGENCOURT_6615160 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5466		
ACCESSION	BM911414		
VERSION	BM911414.1		
	1 GT:19361793		

Db 460 GCGCGCGCGCGCGCGCGCGCGCG 435

RESULT 13
BM911414
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BM911414 1242 bp mRNA linear EST 12-NAR
AGENCOUNT_6615160 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:546
5', mRNA sequence.
BM911414
BM911414.1 GI:19361793
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1242)
NIH-MGC <http://mgc.nci.nih.gov/>;
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LCM1966 row: c column: 19
High quality sequence stop: 350.
Location/Qualifiers
1. .1242
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5466186"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="PH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:
EcoRI; CDNA made by oligo-dT priming. Directionally c
into ECORI/XhoI sites using the following 5' adaptor:
GCGACGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of Californ
Berkeley) using ZAP-CDNA synthesis kit (Stratagene) a
Superscript II RT (Life Technologies). Note: this is
NIH_MGC Library."

47 a 641 c 440 g 55 t 59 others
BASE COUNT
ORIGIN

```
Query Match 4.1%; Score 81.2; DB 14; Length 1242;
Best Local Similarity 43.5%; Pred. No. 0.00025;
Matches 373; Conservative 0; Mismatches 480; Indels 5; Gaps 3;

Oy 43 CCGGCGGACCGCCGCGGCTGCTCCTACGCTAGAGGGCCCTCTCTCAGGTCTTC 102
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 274 CCGGAGGAGTTCGGAGTGGCTGACCTGACGATGATGGGGTCCCGCAGGCGTGC 333
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 103 GTCGGGCGGCTGAGCGGACTGCTGCGAACCTGCTCCCTCGGGGCGCGGCTGTGACC 162
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 334 AGGAGGAGGNNNNNNNNNNCCNGNCCNGNCCNGNCCNGNCCNGNCCNGNCC 393
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 163 G-CGAGCTCAGAGCGGACGCGCAGGAAACGCTGACGTGACCGGATGACGTGAGGC 221
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 394 GCGGCGCGCGGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 453
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 222 GCCAGCGATGCTGACGCGCCCTGAACCTGCGCGCGTCTTCTGGGCTGACGAGCGCTCCAC 281
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 454 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 282 CCGACAGCTCTCTTCTACCGTGTCCCGAAGCGGCTGAGGCTTCTGTCGGCGCCACCGCG 341
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 514 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 573
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 342 TCCAGGAGGAAACAGACGATGACGCGCGGCTATCTACGCGCGCGCGCTGTCCAGGAGC 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 574 GCGCGGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 633
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 402 CGGCGCGGAGCTCGCGCTGATCTTCGGGCGTGTCTGTCGGCGCTGACGCGCGGCGATC 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 634 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 462 CGCTCCGGGCTCGCTTCTGCTGCTACACATGACGCGCGCGCGCGCGCTGCTCGGA 521
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 694 GCGCGGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 522 CCGCAAGCGCGTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 581
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 754 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 582 GAAACAGCGCGCGCGCGCTGGGACGAGCGAGCGAGCGCGCT--TCAGC 639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 814 CNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 640 CACACGCGACCGTATGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 699
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 874 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 700 GTCGCGCGCGGACACGCGAGG--GTTGCGCGCGCGCGCGCGCGCGCGCGCGCG 757
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 934 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 758 AGCGAGCGGCGGCTCGCAAGTACAGGAGCAGTACCTACTGCTGCTGACGCGCGCG 817
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 994 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1053
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 818 GGTTCAGTTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1054 CNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1113
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 878 CCAGCGCGCGCGAGATCG 895
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1114 CNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
BG852363
LOCUS
DEFINITION
1024034A03.y2 C. reinhardtii CC-1690, normalized, Lambda zap II
Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION
BG852363
VERSION
BG852363.1 GI:14233547
KEYWORDS
EST.
```

```
SOURCE
ORGANISM
Chlamydomonas reinhardtii.
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.
REFERENCE
1 (bases 1 to 1281)
Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants: project phase 2
JOURNAL
Unpublished (2000)
COMMENT
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES
source
1..1281
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 2lgr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light, TAP medium in the dark, HS (minimal) medium in
ambient levels of CO2 and HS medium bubbled with 5% CO2.
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with Exassist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT 30 a 697 c 456 g 15 t 83 others
ORIGIN
Query Match 4.1%; Score 81.2; DB 12; Length 1281;
Best Local Similarity 43.1%; Pred. No. 0.00026;
Matches 418; Conservative 0; Mismatches 547; Indels 5; Gaps 3;
Oy 9 CCGGACTTCCTCGAGGACACCGTCCCTACCGCGCGAGCGCGCGCGCGCGCGTCT 68
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 314 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 69 CTTAGCGCTAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCT 128
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 374 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 129 GAACCTGTCTCTTGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 188
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 434 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 189 AACGTGACGTGACGCGGATGCGAGTGCAGCGCGCGCGCGCGCGCGCGCGCG 248
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 494 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 553
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 249 TCGCGCGCGTCTTGGGCTGACGAGGCGTCCCGCGCGCGCGCGCGCGCGCGCGCG 308
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 554 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 613
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 309 GAACAGGTGAGGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 368
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 614 GCGCG--GGCGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 671
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 369 GGCCATCTCACGCGCGCGCGTCTCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCG 428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 672 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 730
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 429 CGCGCGTGTGTCGCCCTGACAGCCCGGGATCCGCTCGGGGCTCGCTCGGTGCGCCGCTCTA 488
Db 731 GGCCTCGCGCCCGCCCGCGGNCNCCGCGGCGNCCCGCGGGCGCGCGCGCGCGCGCG 790
QY 489 CCACATGAGCCCGCCCGAGCGGCTGTGCGAGCCCGCCAGACCCCGGCTCACCACCCACG 548
Db 791 GCGGNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 850
QY 549 GCGCTACCAAGCTGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 608
Db 851 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 910
QY 609 GCGGAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 668
Db 911 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970
QY 669 GCACTTCCCGCTGTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 728
Db 971 GNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1030
QY 729 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 788
Db 1031 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1088
QY 789 GCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 848
Db 1089 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1148
QY 849 CATCGTCAACACGAGGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 908
Db 1149 CGNGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1208
QY 909 GTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 968
Db 1209 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1268
QY 969 AGCAGTGGCG 978
Db 1269 CGCGAGCGG 1278
```

RESULT 15

```
BQ689718
LOCUS BQ689718
DEFINITION AGENCOURT_8344784 NIH_MGC_110 Homo sapiens cdna clone IMAGE:6247838
5', mRNA sequence.
```

```
ACCESSION BQ689718
VERSION BQ689718.1 GI:21815034
```

```
KEYWORDS EST.
```

```
SOURCE human.
```

ORGANISM

```
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
1 (bases 1 to 1051)
```

```
NIH-MGC http://mgc.nci.nih.gov/.
```

```
National Institutes of Health, Mammalian Gene Collection (MGC)
```

```
Unpublished (1999)
```

```
Contact: Robert Strausberg, Ph.D.
```

```
Email: cgapbs-re@mail.nih.gov
```

```
Tissue Procurement: ATCC
```

```
CDNA Library Preparation: Rubin Laboratory
```

```
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

```
DNA Sequencing by: Agencourt Bioscience Corporation
```

```
Clone distribution: MGC clone distribution information can be
```

```
found through the I.M.A.G.E. Consortium/LLNL at:
```

```
http://image.llnl.gov
```

```
plate: LLC#2386 row: 1 column: 15
```

```
High quality sequence start: 2
```

```
High quality sequence stop: 283.
```

```
Location/Qualifiers
```

```
1..1051
```

```
/organism="Homo sapiens"
```

```
/db_xref="taxon:9606"
```

FEATURES

```
source
```

```
/clone="IMAGE:6247838"
/clone_lib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pONB7; Site:1: XhoI;
Site:2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT 63 a 557 c 291 g 69 t
```

```
Query Match 4.0%; Score 80.8; DB 14; Length 1051;
```

```
Best Local Similarity 42.8%; Pred. No. 0.0029;
```

```
Matches 316; Conservative 0; Mismatches 415; Indels 8; Gaps 1;
```

```
QY 7 CACCCCGACTTCCCTCGACACACCGTCCCTTACCGCGCGCGCGCGCGCGCGCGCGCGCG 66
```

```
Db 310 CMCNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 369
```

```
QY 67 CTCCTCAGCCTAGACGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 126
```

```
Db 370 CCCCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 429
```

```
QY 127 TCGAACCTTGTCTCTCTGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186
```

```
Db 430 CCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 489
```

```
QY 187 GGAACCGTGTACGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 246
```

```
Db 490 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 549
```

```
QY 247 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306
```

```
Db 550 GCGCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 609
```

```
QY 307 CGGAACAGGTGTGAGCGTTCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366
```

```
Db 610 CCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 669
```

```
QY 367 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426
```

```
Db 670 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729
```

```
QY 427 TTCGGCGGTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 478
```

```
Db 730 CNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 789
```

```
QY 479 GTGCGGTCTTACACATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 538
```

```
Db 790 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 849
```

```
QY 539 CCACCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 598
```

```
Db 850 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 909
```

```
QY 599 GCTGGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 658
```

```
Db 910 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 969
```

```
QY 659 CGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 718
```

```
Db 970 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1029
```

```
QY 719 CAGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 737
```

```
Db 1030 CGNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1048
```

Search completed: May 10, 2003, 09:52:32

Job time : 2907 secs

THIS PAGE BLANK (USPTO)

[illegible]

```

US-09-103-840A-1/G
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

```
Query Match      4.7%; Score 94.8; DB 4; Length 4411529;
Best Local Similarity 43.5%; Pred. No. 5.2e-09;
Matches 822; Conservative 0; Mismatches 1027; Indels 42; Gaps

QY   75  CTTAGACGGGCGCCCTCTCTCAGGTCTTCTCGGGGACGGTGAGGCAGCTGCTCGCAACT 134
      ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    3936368  CCCGCCGACGCGCGCGGCTTCGCGCTTCGCGCGGCGCGCGCGCGCGCGCTTCGCGC 3939368

OV   135  GGTCCTCTTGGGGCGCGCGGTGTGACCSCGAGCTCTGAGACGGCACGCCAGCAACCGT 194
```

Db 3936308 CTTACACCGTGGCGCGTGGCTAGCCCGGTGCTGTTAGCGCGCGCTTTGCGCGCTTGGC 3936249
QY 195 GCACGTGACCGGATCGAGCTCGAGCGCCAGCGATGCTGACGCCCTTGAAACCTGGCCG 254
Db 3936248 CGCGTGGCGCGCGCACCACTTTGGCGCGCTCAATGCTGGGTCCCGTGGCGCGG 3936189
QY 255 GGTTCGTGGCTGACGAGCGCTGCC---ACCCGACAGCTCTCTTCTACCGTGGCCGAA 311
Db 3936188 GCG 3936129
QY 312 CCAGGTGGACGCTTCTGTCGCGCGCCACCGCTCCACGAGAGAACAGCAATACGCGCGC 371
Db 3936128 GCGCAATCCGGGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 3936069
QY 372 CATCTACCGCGCGCGTCTCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 431
Db 3936068 GCG 3936009
QY 432 CGGTGCTGTCGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 491
Db 3936008 CCGCGCGGTGCG 3935949
QY 492 CATGACG 551
Db 3935948 CGACCG 3935889
QY 552 CTACAGCTGTACTACTGCTGCTGCGGACGAGCAACAGCGCGCGCGCGCGCGCGCG 611
Db 3935888 CGCTGTCG 3935829
QY 612 GAGCAGGACG 671
Db 3935828 CCGCGGTTCG 3935769
QY 672 CTTCCCGGTGTCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 731
Db 3935768 TGGCGCGGACCGCGGATGAAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935709
QY 732 CGCGCGGTGCTGCG 791
Db 3935708 CCGCGCGGTGCG 3935649
QY 792 GTACCTTACTGTCGACGCGCGCGGTTCAGTTTACGCGCGCGCGCGCGCGCGCGCG 851
Db 3935648 CTTTCCCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935592
QY 852 CGTCAACACCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 911
Db 3935591 GGTCCCGTGGCG 3935532
QY 912 CCGCGACCGCAAGATCCACTGGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 971
Db 3935531 CGTCGCGGGAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935472
QY 972 ACTCGGTACGCGCGGTCTACACCTCGCGCAACCTGCGCGCGCGCGCGCGCGCGCG 1031
Db 3935471 GCGCGCGGTACACCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935412
QY 1032 CTTTCACTACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1091
Db 3935411 TCTGGCGGTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935352
QY 1092 CGCAGACGCGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1151
Db 3935351 CGCGCGCGCGGTTCGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935292
QY 1152 CTTCCCATGAGTACGCGTACTGACAGGACACTGGGACGCGCGCGCGCGCGCGCG 1211
Db 3935291 CGTCTTGGGAGAGGCGCGCGGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935232
QY 1212 CGACCTACCGCGCAATGGCTGACTGGGTGGGACTGGTACGCGCGCGCGCGCGCG 1271

Db 3935231 CGCGATACCGCGCGTTCGCCCGCGGTTCGCGCGGTGTTGGCGAGCGCGCGCGCGG 3935172
QY 1272 ATCGATGACGCGCGCGCGGACCAAGCGCTCGCCATCGCGTGGATGAACAACCTGGA 1331
Db 3935171 GCGCGCGGTTCGCCCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935112
QY 1332 CCGCGACGCGAGCTGCCCGACGCGCATCCGACGCGCTACACGCGGAGAACTCGATG 1391
Db 3935111 TGGCG 3935071
QY 1392 CGCGGAGTGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1451
Db 3935070 CTGCGCGGTGGCTAGCGCGGTGCTGTTAGCGCGCGCTTTGCGCGCGCTTG 3935013
QY 1452 GCGAGCGTGGAGAACTAGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1511
Db 3935012 GCG 3934953
QY 1512 CAGCGCGGTCTGCCATGGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1571
Db 3934952 CCG 3934905
QY 1572 CACCGCGAGAACTGCGCGCATCTCGGTGGCGCGCTCCCGCGCGCGCGCGCGCG 1631
Db 3934904 CACCG 3934845
QY 1632 CATCGCGAAGTACGCGAGCGCGCTGTAGTCCGACCGAGGACCGCTCCGCGCGG 1691
Db 3934844 ATCG 3934788
QY 1692 CTCGCTGCG 1751
Db 3934787 GTCG 3934728
QY 1752 CCGTGGCATCTCTGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1811
Db 3934727 CCG 3934668
QY 1812 GCTCTCCCGAGGTGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1871
Db 3934667 GCGCGCGCGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934608
QY 1872 CCG 1931
Db 3934607 CCGCAACCG 3934548
QY 1932 GCACACCGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1962
Db 3934547 CACCG 3934517

RESULT 4
US-08-403-852D-1
; Sequence 1, Application US/08403852D
; Patent No. 5891695
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanc, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Crecy-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved In The
; Biosynthesis Of Streptogramins, Nucleotide Sequences
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington

STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852D
FILING DATE: 10-MAY-1995

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441

ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146

REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5392 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: cDNA
HYPOTHETICAL: NO

ANTI-SENSE: NO
ORIGINAL SOURCE:

ORGANISM: S.pristinaespiralis
US-08-403-852D-1

Query Match 4.6%; Score 91.4; DB 2; Length 5392;
Best Local Similarity 42.1%; Pred. No. 1.8e-08;
Matches 652; Conservative 0; Mismatches 891; Indels 6; Gaps 2;

Qy 382 CGCGCCGTGCTCCAGGAGCGCGCGGAGCAGCTCGCCCTGATCTTCGGCGGTGCTGTG 441
Db 2748 CGGCACGATCAACCGCGCGCTGCGGGCGCGGAGCGCGCGGATTCGACGCGCTCTGAT 2807
Qy 442 CCGCCTGCGAGCGCGGATCGCTCGGGCTCGCTCGCGCTGCTACACATGACGCGCC 501
Db 2808 CGACGACGCGCGCGCGCGCTCCAGGCGCGGTTCGAGACGAGGAC--CTGACCGCC 2865
Qy 502 CCCAGCGCTGGCTGTGCGACCCCAAGCGCGGTGTCACCCAGCGCGCTTACAGGTG 561
Db 2866 GCGTGGCGCGCGTCAACGAGCAGATCGGCTGATCAGCGCGCGCTCCGCGCGACGAG 2925
Qy 562 TACTACCTGCACTCCGACCAACAGCGCGCGCGCGCTGATGCGCGCTGCGCGCGAGCAG 621
Db 2926 GCCCGCTACCACTGCTCCGATCACCGCTCGCTCGACCACTCGCGCGCGCGCGCG 2985
Qy 622 GACGCGCTGCGCTTCAACGACGAGCGCGCGCTGATGCGCGCTGCGCGCGAGCTTCCCG 681
Db 2986 GGCTGGCTCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3045
Qy 682 TGGTCCGGGTGCGGCTGCTCGGACCGCGGACGAGCGGTTTCGGCGCGCGCGCGGTG 741
Db 3046 GTCGTCCGCGCGCTGTGGAGAGCTTCGACGACGAGCGCTTCGTCACGAGCGCGCGCG 3105
Qy 742 GTCGGCTCGGACCGCGCGCGCGCGCGCTCGCGAGTACGAGGAGCAGTACCTCTAC 801
Db 3106 GGCCTGTACTGGCGCTGCGCGCGCGCTCCACCACTCGACCACTCGAGGCGAGCAGTTCG 3165
Qy 802 TGGTGCACCGAGCGGCTTCACTGCTTCAACCGCGCGCGCGCGCGCTGATCGTCAACAC 861
Db 3166 GTGGCGCGCGCGCTTCACTGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3225

Qy 862 GACGCTCG 921
Db 3226 GCGCGCGCGCTCG 3285
Qy 922 AAGATCCACTGGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 981
Db 3286 TCGGTGAAGCAGCAGCG 3345
Qy 982 GCGCGCTTCTACACTCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1041
Db 3346 GAATGCG 3405
Qy 1042 CCGAACACCGCGCTCG 1101
Db 3406 CCGGTCTTGGCGCGCTCG 3465
Qy 1102 GGGACACCGCGCTCG 1161
Db 3466 CTGCGCGAAGCGCTGGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3525
Qy 1162 ACGTACGCGCTACTGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1221
Db 3526 CGACCGCTCG 3581
Qy 1222 CCGCAATGGCTCGACTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 1281
Db 3582 GAGTCCGTGACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3641
Qy 1282 GCGCGCGAGACCAAGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCG 1341
Db 3642 GACGCGCTGCTGCGCGAGGACCGCGCGCTGACGCGCTGCGCGCTGCGCGCTGCG 3701
Qy 1342 GACGTCCCG 1401
Db 3702 GCGCAGGTCCACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3761
Qy 1402 CGGTGCG 1461
Db 3762 GTCCGCGACGAGATCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCG 3821
Qy 1462 ACGAATACGCTCACCG 1521
Db 3822 TCCTGCGCGCTGCTCGCTTCCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3881
Qy 1522 CTGCATGGAAGCG 1581
Db 3882 AGCGCTACGAGACCGCGCTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941
Qy 1582 AACGTGCGCATCTCGGTGGCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1641
Db 3942 GACGAGGCGCTGATGTTGCGGTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4001
Qy 1642 TACGAGCAGACCTGTAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1701
Db 4002 ATCAGGTGCGCGCGCGCGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4061
Qy 1702 CCCTACTCG 1761
Db 4062 CCCTACTCG 4121
Qy 1762 CTCGTGACACCG 1821
Db 4122 GTGGCGCTGGACACCGCTGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4181
Qy 1822 CAGGTCCACTTCG 1881
Db 4182 CTGCTACCG 4241
Qy 1882 CACTTTCACCG 1930
Db 4242 GACGCGCATCAAGCTCGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4290

QY 1222 CGCAATGGCTCGACTGGGGCTGGGACTGGTACGGGGCGCTCACCTGGCCATCGATCGAC 1281
DB 3582 GAGTCGTCGACGGAGGCGACCCGACAAAGATCGCGACAGATCAGTGACACCGTCTCTC 3641
QY 1282 GCGCCGAGACCAAGCGCTCGCCATCGGTGGTGATGACAACTGGAAGTACGCCGACGC 1341
DB 3642 GACGCGCTGCTCGCGGAGGACCCCGCTCACCGCTCGGGGTGAGACCGCTGATCACACC 3701
QY 1342 GACGTCGCCACCGACGATCGGACGGCTACAAAGGCGAGACTCGATCGTCGGCGAGTG 1401
DB 3702 GCGGAGTCCACATCGCGGCGAGGTACCAACAGGCTACGCCCGCATCGCCCACTG 3761
QY 1402 CGGCTGCGCCGACACCTGGCGGTGGTACACCTCTTGGACACCCCGTGGCAGCGTG 1461
DB 3762 GTCCCGGACACGATCTGGCCATCGGTACGACTCGTCCGCAAGGGCTTCGACGGCGCC 3821
QY 1462 AGCACTAGTACCGCCACCAACACACTCTCCCGACGGAGCGGTGAGCGAGCGCCGTC 1521
DB 3822 TCCTCGGGGTCTCGGTCTCCATCGGCGCGGACGTCCCGGACATCGCCAGGGGTGAC 3881
QY 1522 CTGCCATGGAAGGACGCGATACGAGATCGAGCTCGACATCGCTGGGACACGCGACG 1581
DB 3882 AGCGCTACGAGACCGCGTTCGAGGGCGAGGACGAGCTCGACGAGCGGCGCCGCG 3941
QY 1582 AACGTGCGCATCTCGGTGGCGGTCTCCCGGACGGAACCCGCGACACGAACATCGGCAAG 1641
DB 3942 GACCAAGGCGCTGATCTCGGTACGCCACCGACGAGACCCCGCTGCTGCTGCGCTGCC 4001
QY 1642 TACGAGGACAGCTGATGCTGACGAGGAGCGCTCCGACCTCGCGGGTACTCGTCCGC 1701
DB 4002 ATCGAGCTCGCCACCGCTCTCGCGCGCGGTCTACCGAGGTCCGCAAGGACGGACCGTC 4061
QY 1702 CCTACTCGGAGCGCGCGCCCGCATCGACATCGGCGCGCGATCGTGCACCTGGCGATC 1761
DB 4062 CCTACTCGGAGCG 4121
QY 1762 CTCGTCGACACCCAGAGCGGTTCGAGTCTTCGTCACCGCGCGCGCGCGCGCGCGCGCG 1821
DB 4122 GTGCGCTGGACACCGTCTGCTCTCTCCAGACGCGCGCGCGCGCGCGCGCGCGCGCG 4181
QY 1822 CAGGTCCACTTCGCGAGGGGACACGGAATCTGCTCTACCGCGCGCGCGCGCGCGCGCG 1881
DB 4182 CTGCTACCCCGGATCGCGGAGCACCTCGTTCGAGACAGTCTCTCGCGCGCGCGCGCG 4241
QY 1882 CACTTCACCGGATCGTCTCGCGGAGATTCGCCAGGCGATCTAGCGGA 1930
DB 4242 GACGCGATCAAGCTCGAGACGACAACTACCGCGTCTGTTGTTCAACCCGA 4290

RESULT 7
US-08-804-227C-7
Sequence 7, Application US/08804227C
Patent No. 5876991

GENERAL INFORMATION:
APPLICANT: DeHoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7

Query Match 4.3%; Score 85.6; DB 2; Length 44377;
Best Local Similarity 42.1%; Pred. No. 2e-07; Indels 30; Gaps 3;
Matches 727; Conservative 0; Mismatches 969;
QY 69 CCTCAGCCTAGACGGGCGCCCTCTCTCGAGGTCTTCTCGGGGACGCTGAGGCGACTGCGTCT 128
DB 26238 CGGACCCGGTTGAGGACGCGGCGCGCTGTCGGCGAGAATTCACGAGCTGATCGC 26297
QY 129 GAACCTGTCTCTCTGGGGCGGGGTGTGACCCGAGCCTCGAGAGCGGACGCCAGG 188
DB 26298 GGCCCTTCGACGACTCTGTCGAAGGCGCCACCGCTCTGCTGCTGCGGGGACCGCGCG 26357
QY 189 AACCTGCACGTGACCGGATCGAGTCTGAGCGCGCCAGCGATGCTGACGCCCTGAACC 248
DB 26358 CACCTCCGACAGGTCTGTTCTGCTTCCCGCGCAGGCTCGCAGTGGCCGAGATGCG 26417
QY 249 TGCCGCGCTTCTGGGCTGACGAGCGCTCCACCGCGACAGCTCTCTTCTACCGCTGCC 308
DB 26418 CGAGCGGCTGTCGCGCGCTCCAGCGGCTCGGCTCTCTTCTGAGAGACCGCGCGCTG 26477
QY 309 GAACGAGGTGGAGCTTCTGTCGCGCCCGCCAGCGCTCCAGAGAGGAAACAGCAATCAC 368
DB 26478 CGACCTCGCTCGCGCGCCCGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 26537
QY 369 GGCATCTCACGCGCGCGCTGCTCCAGGAGCGCGCGCGCGGAGCACTCGCCCTGATCTT 428
DB 26538 CGGCGCGCTCTGCTGACCGGGTTCGAGTGGTGGAGCGCGCTGCTGCTGCTGCTGCT 26597
QY 429 CGGCGGTCTGTGGCGCTGACGCGCGGCGATTCGCTCGGGGCTCGTCCGTCGCTGCTA 488
DB 26598 CTGCTCGCGAGAGTGGCGTTCGCTGGGCGTTCGAACCGCGCGGCTGCTGCTGCT 26657
QY 489 CCACATGACGCGCGCGCGCTGCTGCGAGCGCTGCTGCGAGCGCGCGCGCGCGCGCG 548
DB 26658 CCAGGCGAGATCGCGCGCGCTAGTCTGCGCGCGCGCTGAGCTGAGGAGCGCGCGCG 26717
QY 549 GCGCTACGAGTGTACTACCTGCACTCCGACGACGAACACCGCGCGCGCGCGCGCG 608
DB 26718 CATGCTCG 26777

Qy	609	CGGAGCAGCAGCAGCGGCGTTCGCGATTCACGCACACGGCACAGGTGATGCGCTGCGGCC	668
Db	26778	CGTGAACCTGTCGGAACGCGACCTCGGTCCCGCTGAGCGCTGAGCGCACCGCGCTCGC	26837
Qy	669	CGACTTCCCGCTGGTTCGCGGTGGCGGTGCTCGGCACCGCGAACACGGCAGGGTT---	725
Db	26838	CGTCCGCCCTCAACGGCCCCGAGACTCGCGCGCTCTCCGGGACCCCGACGCCCTGCG	26897
Qy	726	-----CGCGCCCGCGGGTCTCGCGCTCGCGACCCAGCCAGCCAGCGCGGT	773
Db	26898	GGAGCTGTGTCGCCGAACCTCGGTGCGGAGGGGTGCACGCCGCCCATCTCCGCGCTCGA	26957
Qy	774	CCGCAAGTACAGAGCAGTACTCTACTGTGTGACCGACGGCGGGTTCAGTTCACCGC	833
Db	26958	CACCGCGGGCACTCGCGCGAGTGCACACGCTGGAGGCCACACTGCGGAAGTGTCTCGC	27017
Qy	834	CCTGCCCGAACCCGCTATCGTCAACACCGAGGTGCGCGCCGCCACCGCCGCCGAGAT	893
Db	27018	GCCGCTCGCGCCCGCCACTCCGACATCCCGTTCTACTGACGGTCAACCGAGGACTGAT	27077
Qy	894	CGAAGCGCGAGTGGTTTCGCGACCCCAAGATCCACTGGACACCGCCCCCGGAGATG	953
Db	27078	CGACACCGCGAGTGGACGCGCACTACTGTGTACCGCAACATGCGCAGCGGTGGAGTT	27137
Qy	954	GSTCTCGTCACTCGGACGACTGCGGTACGCGCGGTTCTACACTTCGCGAACCTTCGCG	1013
Db	27138	CGAGAGGCCACCCGCGCCCTGATGCGCGACGGCCACGACGTCTCTTGGAGTCCGAGCC	27197
Qy	1014	CTGGACACTTCCGCGCAACTTCGACTACCCGAAACACGCCCTCGCGGCGCATCGAGTGCC	1073
Db	27198	GCACCCATGCTGGCGCTCTCCCTCCAGGACAGCATCAGCAGCGCGGTTCCCGCGCGC	27257
Qy	1074	CGACTGTTGAGATCACCGACAGCGGACACGCCACTGGGTGCTCGCGCGCCAGCAT	1133
Db	27258	CGTCTCGGACCCGTGGCGCGCGCACAGGGCGGCCGCGCTGGCTGGCGCTCGCCCTCG	27317
Qy	1134	GGAGCGCTACGGCATCGGCTTCCCATGACGTAGCCTACTTGACAGCGACCTGGGACGG	1193
Db	27318	CCGCGCTTACACCCAGCGCTTGGAGATCGACCGCAGGCGCATCTTCGGGCCCGACTCAG	27377
Qy	1194	CGAGCAGTTCCACGCCGACGACCTCACCCCGCAATGGCTCCACTGGGCTGGGACTGGTA	1253
Db	27378	CCAGTGGAACTGCCACGTACCCCTTCCACGCGGAGGCTTACTGGTACACCCCGGCCA	27437
Qy	1254	CGCGGCGTCACTTGGCATCGATGACGCGCGCGGACCAAGGCCCTTCGCCATCGCGTG	1313
Db	27438	CCGCGGTGACACCCCGCTCCCTCGGTCTGGAGCGCGCTCGACACCGCTGCTGGGCGA	27497
Qy	1314	GATGAACAATGGAAGTACGCCGACGCGAGCTCCCGACCGACGATCCGACGCGCTACAA	1373
Db	27498	CGGCGTGGAACTGCGGAGTCCGGTGACCGGATGTACACCGACGCGGTGGCGCGGACAC	27557
Qy	1374	CGGCGAANAATCGATCTGTCGCGAGCTCGGCTCGCCCGACAGCTGCGGCGCTGGTACAC	1433
Db	27558	CACCCGTGGTGGCGACACGCGCTGCTGGGGTGCCTGCTGCGCGCGCGCGCTT	27617
Qy	1434	CCTCTGAGCACCCCGTGGCAGCGTTCAGAACTACGCGGACACCCACACTGCC	1493
Db	27618	CG-----CCGACCTGGCGCTCTGGGCGCGGCCGACGCGCGACCGCGCGCTCGA	27668
Qy	1494	CGACCGGACCTGACGCGACGCGCTCTGCCATGGAAACGACCGCATACGAGATCGA	1553
Db	27669	GGACTCACCTGGCGGCCCTTGTGCTGCCCGCTTCGCGGGGTGTCCGGCTGCGGCT	27728
Qy	1554	GCTCGACATCGCTTGGACACCGGAGAACTGCGGATCTCGGTGGGCGCGCTCCCGCA	1613
Db	27729	GAACTGGCGGCCCGGSCACCGACGACGCGCCGCGCTTCGCGTGCACGCCCGCGCGA	27788
Qy	1614	CGGAACCGGCACACGAACATTCGCGAAGTACGAGACAGACTGTACGTGCGACGAGACC	1673
Db	27789	GGGCGCCACGACTTGGACCTT-----GCACCGGAGGGGTGCTCAACCGCGCAGGACAC	27842

Qy	1674	CTCGACACTCCGCGGTACTGCTCGGCCCTTACTTCGGAGCGCGCCCCCATCGACC	1733
Db	27843	GCGCAGCGCGGAGCGCTTCGGCGGCACCCC CGCGCGCGCAACAATGGACAT	27902
Qy	1734	CGGCGCCGATCCGTTGCACTTCGCGCATCTCGTCGACACCCAGAGC	1779
Db	27903	CGCGACTTCTACAGCGCTTCTCGGAACCTCGGTTACGGCTACGC	27948

RESULT 8
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rosteck, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804.198

[illegible]

```

? FILLING DATE:
? ATTORNEY/AGENT INFORMATION: 435
? NAME: CANTRELL, PAUL R.
? REGISTRATION NUMBER: 36,470
? REFERENCE/DOCKET NUMBER: P9113
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 317-276-3885
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 44377 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 350..14002
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 14046..20036
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 20110..31284
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 31329..36071
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 36155..41830
? US-08-804-198-1
?
? Query Match 4.38; Score 85.6; DB 2; Length 44377;
? Best Local Similarity 42.1; Pred. No. 2e-07;
? Matches 72; Conservative 0; Mismatches 969; Indels 30; Gaps
?
? 69 CCTCAGCCCTAGACGGGCCCTCTTCGTGCGGGACGGTGAGGGCACTGCCTC 128

```

Db 26238 CCGCACCCGGTTCGACAGCGGGCCCGCTGCTGCGCGAGAACTTCACGGAGCTGATCGC 26297
QY 129 GAACTTGGTCTCTCTGGGGCGGGCGGTGTGACCCGAGACCTTCGAGACGGCACGGCCAGG 188
Db 26298 GGGCTTCGAGGACCTCTGTCGAGGGCGACCGACCGCTCTGCTGCGGGGACCGCCCG 26357
QY 189 AACGTTGACGATGACCGGATGACGATCGAGGCGGCCACGCGATGTGACGCCCTTGAACC 248
Db 26358 CACCTCCGACAGGTGCTGCTGCTTCCCGGCGCAGGCTCGCAGTGGCCCGGAGATGC 26417
QY 249 TCGCGCGTCTTGGGCTGACGAGCGTCCACCCGACAGCTCTCTTCTACCGTGGCC 308
Db 26418 CGAGGGCTCTGGGCGGCTCCAGCGGCTCCGGCTCTTCTTCGAGACCGCCCGCGCTG 26477
QY 309 GAAACAGGTGACGCTTCTGTCGCGCCACCCCTCCACGAGAGGAACAGCAATGACGC 368
Db 26478 CGACTTCGGCTTCGGCCCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26537
QY 369 GGCATCTCACCGCGCGCTGCTCCAGGAGCGCGCGCGGAGCACTCGCCCTGATCTT 428
Db 26538 CGCGCGCGCTCTGCTGACCGGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26597
QY 429 CGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 488
Db 26598 CTGCTCGCGGAGACGTGGCGTTCGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26657
QY 489 CCACATGACCGCCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 548
Db 26658 CCAGGCGAGATCGCGCGCGCTACGTCGCGCGCGCTGACGCTGACGCTGACGCTGAC 26717
QY 549 CGCTTACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 608
Db 26718 CATGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26777
QY 609 CGGAGCAGCAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 668
Db 26778 CGTGACCTTGTGCGAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26837
QY 669 CGACTTCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 725
Db 26838 CGTGGCGCGCTCAAGCGCGCGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26897
QY 726 -----CGGGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 773
Db 26898 GGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26957
QY 774 CGCAAGTACAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 833
Db 26958 CACCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27017
QY 834 CTTGCGCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 893
Db 27018 GCGCGTGGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27077
QY 894 CGAGAACCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 953
Db 27078 CGACACCGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27137
QY 954 GGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1013
Db 27138 CGAGCAGGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27197
QY 1014 CTTGAGACTTCGCGCGAACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1073
Db 27198 GCACCCCATGCTGGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27257
QY 1074 CGACTTGTGAGATCACCGCAGACCGGAGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1133
Db 27258 CGTCTTCGCGACCTTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27317
QY 1134 GGACGCTACCGCATCGGCTCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1193
Db 27318 CCGCGCTACACCCCGGCTGGAGATCGACCGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27377

QY 1194 CGAGCAGTTCCAGCGCGACGACCTCACCCGCAATGGCTCGACTGGGGCTGGGACTGGTA 1253
Db 27378 CCAGGTGGAACCTGCCACAGTACCCCTTCAGCGCGAGCGCTACTGTGTACACCCCGGCCA 27437
QY 1254 CGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1313
Db 27438 CGCGGCTGACGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27497
QY 1314 GATGAACAACTGGAAGTACGCGCGACGCGACGCTCCACACCGGAGCATCCGAGGCTACAA 1373
Db 27498 CGGCTGGAACCTGCGGAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27557
QY 1374 CGGCGAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1433
Db 27558 CACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27617
QY 1434 CTTCTGAGACACCGCGTGGAGCGCTGACGAACTGACGTACCGCGGACACACACTCC 1493
Db 27618 CG-----CCGACTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27668
QY 1494 CGACCGGACCGTCCGCGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1553
Db 27669 GGAGCTACCTTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27728
QY 1554 GCTCGACATCGCTGGGACCGCGACGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1613
Db 27729 GAACTGCG 27788
QY 1614 CGGACCG 1673
Db 27789 GGGCG 27842
QY 1674 CTCCGACCTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1733
Db 27843 GCG 27902
QY 1734 CGGCG 1779
Db 27903 CGGCG 27948

RESULT 9

US-07-945-283-1/c
; Sequence 1, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutant.s
; TITLE OF INVENTION: Involving The EP0 and LLT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Peoria
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
; FILING DATE: 19920911
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ribando, Curtis P
; REGISTRATION NUMBER: 27976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 309-685-4011 ext.513

```
TELEFAX: 309-685-4128
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8438 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Pseudorabies virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 622..6495
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1099, "g")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1267, "t")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1381, "c")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1566, "c")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(7010, "g")
;
; US-07-945-283-1
Query Match 4.2%; Score 84; DB 1; Length 8438;
Best Local Similarity 43.0%; Pred. No. 3.7e-07;
Matches 640; Conservative 0; Mismatches 830; Indels 18; Gaps 4;

Qy 329 TCGGCGCCACCGTCCAGAGAGAACACGAAATGACGCGGCCATCTACGCGCGCGCG 388
Db 5549 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5490

Qy 389 TGCTCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 448
Db 5489 CTCTCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5430

Qy 449 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 508
Db 5429 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5370

Qy 509 GCTGGCTCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 568
Db 5369 ACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5310

Qy 569 TGCACCTCGGACGAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 628
Db 5309 CCGACCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5250

Qy 629 TCGCGCTTCACGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 688
Db 5249 TCTGGAGAGAGGACGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5190

Qy 689 GGTGGCGGGTCTGGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 748
Db 5189 TGCGCGCTTCATCCGCGAGATGGGGACTCGAGGAGACGACGAGCGCGCGCG 5130

Qy 749 TCGCGACCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 808
Db 5129 TCATCTAGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5070

Qy 809 CCGAGCGGGGTTACGTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 868
Db 5069 CCGCGGACCGCGCTTCACGAGTTCTCCGAGCGCGCGCGCGCGCGCGCGCG 5010

Qy 869 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 928
```

RESULT 10

```
US-08-951-742-1
; Sequence 1, Application US/08951742
; Patent No. 6127144
; GENERAL INFORMATION:
; APPLICANT: Bartfield, Daniel
; APPLICANT: Michael J. Butler
; APPLICANT: Dany Hadary
; APPLICANT: David Jenish
```


RESULT 11
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatenIn ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 4.1%; Score 83; DB 4; Length 4403765;
Best Local Similarity 42.6%; Pred. No. 6.6e-07;
Matches 774; Conservative 1; Mismatches 1029; Indels 12; Gaps 4;

Qy 86 CCTCTCTGAGGCTCTCTCGGGGACGGTGGAGGACACTGCGTCGAACTGGTTCCTCTGG 145
Db 1631220 CCTCTGTTGGCGAGGACGCCCTTGGCGCGGTGTCGCGCGCGCCAGTCCCGCG 1631279
Qy 146 GGGCGCGGGTGTACCGGAGCCTTCGAGAGCGGACGCGCAGGAAACCGTGCACGTGACG 205
Db 1631280 ATGCCACCGTTGCCCGCGTGGCGCGGTGCGCGCTTACCGCGCTTACCGCGCG 1631339
Qy 206 CGATCAGCTCGAGGCGCGGACGATGTCAGCGCCCTGAACTGCGCGCGTTCCTGGGCT 265
Db 1631340 CCGGGCGCGCGTTCGCGCGCTGGCAGCGCGTGGCGCGGTGACCGCGGTTCGCCCG 1631399
Qy 266 GACGAGCGCTCCACCCCGACAGCTCTCTCTACCGCTGCGCGGACGAGGTGACGCT 325
Db 1631400 TCGTGMGGATGCTCCGCGCGCGCCCGCTTGGCTGCGGTGGAGCGGTGCCCGCGG 1631459
Qy 326 TCGTCCGCGCGACCGTCCAGAGAGGAAACGACGATGACGCGCGCGCATCTACGCGCG 385
Db 1631460 CCGCGCGCACCGGCTTGGCGCGCTTGGCGCGCTTGGCGCGCTTGGCGCGCGCG 1631519
Qy 386 CCGTCTCTCAGGAGCGCGCGGACGACTCGCCCTGATCTTGGCGGTGCTGTGCGCG 445
Db 1631520 GCGAAGCGCCCTGCTCCCTGCGCGCTTGGCGCGGTGGCGCGGTGACCGGAGCACCGCG 1631579
Qy 446 CTGACGCGCGGACATCCGCTCGGGCTCGCTCGGCTGTACACATGACGCGCGCGCA 505
Db 1631580 TCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1631639
Qy 506 GGGGTGCTGTGACCG 565
Db 1631640 CCG 1631699
Qy 566 ACCTGCTCTCAGGAGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 625
Db 1631700 GCGGACCG 1631756
Qy 626 GCGTGCCTTACGACCG 685
Db 1631757 AGCGCGCGTGGCG 1631816
Qy 686 CCGGTGCGCGGTGCTGCGGACCGCGGAAACAGGCGAGGCTTGGCGCGCGCGCGCGCG 745

Db 1631817 CCGCGCGCAC---CGTTGCCCGCGGGTTACCGTTTGGCGCGGTTTACCAAGGTTGGTG 1631873
Qy 746 CGCTCGGACCGACCGGAGCGGTCGCGCAAGTACCAAGGAGTAGTACTCTACTGTT 805
Db 1631874 GCGTTGGCACTCATGCCACCAAGCGCGCTCGCGCGCGCGCGCGCGCGCGCG 1631933
Qy 806 CGACCGAGCGGGGTTACGTTTACCGCGCTGCCGAGACCGCGCTCATCGTCAACACCGAG 865
Db 1631934 CCGCGCTGCGCGGTTGCGCGCATTTGCGCGCTGCGCGCGCTGCGCGCGCGCGCG 1631993
Qy 866 CTCG 925
Db 1631994 GAGTTGCGCGCGTTCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632053
Qy 926 TCCACTGGGACACCGCGCGCGGAGATGGGTCTGGTCTATCGGACGACTGCGGTACGCG 985
Db 1632054 GCGCGCTTGGCGCGTGGGCTAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632113
Qy 986 CGTTCTACACCTTCGCGCGCACTTGGCGACTGGACACTTCGC---CGCAACTTCGACTACC 1042
Db 1632114 GGGTTACCGCGTTACCGCGGTGACCGCGTTACCATCGCGGAGGCGAGTTGCGGTG 1632173
Qy 1043 CGAACACCGCTCGCGCGCATCGAGTGCCTGCGCGCGCGCGCGCGCGCGCGCGCG 1102
Db 1632174 GCGCGCTTGGCGCGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632233
Qy 1103 GGACAGCGCACTGGGTGCTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1162
Db 1632234 ACACCTTGGATTCGTTCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632293
Qy 1163 CTTAGCGCTTCTGGACAGCGCACTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1222
Db 1632294 CCGCGCTTACCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632353
Qy 1223 CGCAATGCTGACTTGGGCTGGGACTGTAAGCGCGCGCGCGCGCGCGCGCGCGCG 1282
Db 1632354 GCGCTGTTGGCGTGGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632413
Qy 1283 CGCGCGAGCAAGCGCTGCCATCGCTGCGTGGATGAACACTGGAAGTACGCGCGCG 1342
Db 1632414 CCG 1632470
Qy 1343 AGTCTCCCGACGACATCCGACGCGTACACGCGCGAGAACTCGATCGTCGCGCGAGTGC 1402
Db 1632471 CCG 1632530
Qy 1403 GGTGCG 1462
Db 1632531 GCGAGGAACCG 1632590
Qy 1463 CGAAGTCTCAGCG 1522
Db 1632591 CCAACAAGCG 1632650
Qy 1523 TGCCATGGAAGCG 1582
Db 1632651 GCGTTCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632710
Qy 1583 AGTGGCGATCTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1642
Db 1632711 GCACCGTCTGACCGCGCGGTGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632770
Qy 1643 ACGGAGGACGCTGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1702
Db 1632771 CGGTACCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632830
Qy 1703 CCTACTCGGAGCG 1762
Db 1632831 TTGCAACCGTTCG 1632890
Qy 1763 TCGTCAACCG 1822

Db 926 CAGCTCGCGCGCATGTGCTGCAGACGGTCTCAACCCGTCCTCCGCGACAGATCTGGTAC 985
QY 1510 GGACGCGCGCTCTGCTGATGAACGCGCATACGAGATCGAGCTCGACATCGCTGG 1569
Db 986 CAGGCGCAACTGTGACAGGAGCTGCGCTTCGAGGGCGCTGGAAGACTGGCAGGACTGG 1045
QY 1570 GACACGCGGACGAACTCGGCACTCTCGGTGGGCGCTCCCGGCGGAAACCCGCGCACAG 1629
Db 1046 GTCCGCGGAAACGACGCGCGCTACCACTCTCGGCGACACCGCGCGAGGTCCAGGACCA 1105
QY 1630 AACATCGGAAGTACGACGACAGCTGTAGCTCGAGCGGAGGACCTCCGACCTCGCGGG 1689
Db 1106 TGCGTGAAGTGGCGCGCGCGCGGCGGAGAGCCGCTGGGGCGCTCGTCGACCGGGC 1165
QY 1690 TACTCGCTCGCCCTTACTCGGAGCGCGCGCCCATCGACCCCGCGCGCGATCCGTG 1749
Db 1166 GAGCTGATCTCTTCTTCAGAGCGCGCGCTACTAGACTCCGCTGGCGCGCGCGCG 1225
QY 1750 CACCTCGGATCTCTGTCGACACCCAGAGGCTCGGTGAGGCTTCTGTCACCGCGCGCACAC 1809
Db 1226 GAGATCTTCAGCAAGTACGTGCGCGCGGACACCCAGGCGCTGTCGACGCGCGCGCACCC 1285
QY 1810 GTGCTCTCCAGCAGCTCCACTTCGCGGAGGCGGACACGGAATCTCGCTTACACCGAC 1869
Db 1286 GACCTGTCCGACACCGCGGCGAACGCTCCGCGGAGAACGGAACCGCTCTACACGGCC 1345
QY 1870 GCGCGCGCGCACATTCACCGGCGATCGTCGCGG 1904
Db 1346 GTCAGTGCACCGACGCGCAAGTGGCGCGCGCAACTG 1380

RESULT 14

US-08-387-942C-1
; Sequence 1, Application US/08387942C
; Patent No. 5939289
; GENERAL INFORMATION:
; APPLICANT: ERTESVAG, HELGA
; APPLICANT: VALLA, SVEIN
; APPLICANT: SKJAK-BRAEK, GUDMUND
; APPLICANT: LARSEN, BJORN
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES
; TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: P.O. BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/387,942C
; FILING DATE: 09-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1809-106P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12588 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:
; ORGANISM: Azotobacter vinelandii
; STRAIN: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 290..1951
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2227..6438
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6702..9695
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9973..12588
US-08-387-942C-1

Query Match 4.0%; Score 80.4; DB 2; Length 12588;
Best Local Similarity 42.3%; Pred. No. 1.7e-06;
Matches 508; Conservative 0; Mismatches 691; Indels 3; Gaps 1;

QY 705 CACCGCGAAACACGGCAGGTTGGCGCGCGGGTCTGCGCTCGCGACCGCGGAC 764
Db 6875 CATCAAGACACAGCTCCATATCGTCGCGCGGGGATGGGGAGACGGTCTCAAGTGGT 6934
QY 765 CGACGCGCTCCGCAAGTACACGAGCAGTACTCTACTGTGTGACGACGCGGGGTTCAC 824
Db 6935 CGACGCGCTGGGATCAGGACGTCACCGGATCGTCGCTCGCGCTACGGCGAGGAGACCAG 6994
QY 825 GTTACCGCGCTCGCGGACCGCGCTCATCGTCAACACCGAGGTCGCGCGCGCACAGGCC 884
Db 6995 CAACCTTCGGCATGAGCGACCTGACCTCGACGCGCAACCGCACACGCGCGCAAGGT 7054
QY 885 CGCGGAGATCGAGACGCGCGAGTGGTTCGCGGACCCCAAGATCCACTGGGACACCGCCG 944
Db 7055 CGACGCGCTGTTCAACGGCTACATTCGCGCGGAGGACGGCGCGGACGTGACCT 7114
QY 945 CG---GAGAATGGTCTGCTGCTCATCGGACGACTGCGGTACGCGCGCTTCTACACCTCGCC 1001
Db 7115 GGAGCGGTGGAATCCGTGAAATGTCGGTTACGGTTTCGATCCGACGAGCAGACCAT 7174
QY 1002 GAACCTGCGGACTGGACACTTCGCGCGCACTTCGACTACCGACCGACCGCTCGCGCG 1061
Db 7175 CAACCTTGACGATCCGCGCAGCGTGGCGCCACGACAAAGCCCTCGACGGCTTCGCGCG 7234
QY 1062 CATCGAGTCCCGGACCTGTTGAGATCACCGCAGACGCGGACGACGCGCACTGGGTGT 1121
Db 7235 TTTCCAGATCGCGGGGTGTTGAGAACAACTCTCGTACAAACGACCGCGCGCTT 7294
QY 1122 CGCGCGCAGATGGAGCGCTACGGATCGGCTCCCGCTAGCGTACGCTACTGGACAGG 1181
Db 7295 CAACATCGTCAACGACCAACGACTTCGCTCTGAGCAACAACGCTGCGCTACGCGCAACGG 7354
QY 1182 CACCTGGGACGCGGAGCAGTTCACCGCGCAGCACTCCCGCGCAATGCTCGACCTGGG 1241
Db 7355 CGCGCGCGCGCTGGTGGTGCAGCGGGCTGTCGAGCTGGCGCACCCCTACGACATCT 7414
QY 1242 CTGGGACTGTACGCGCGCGCTCACCTGGCCATCGATCGACGCGCGCGGAGACCAAGCGCT 1301
Db 7415 GATCGACGCGCGCGCTACTACGACAAACGCGCTGGAAGCGGTGCGAGATCAAGATGCCCA 7474
QY 1302 CGCCATCGGTGGATGAACAACCTGGAAGTACGCGCGCAGCGAGTCCCGACCGACCATC 1361
Db 7475 CGACGTACCCCTGCAGACGCGGAGATCTACGCGCAACGCGCTATACGGGTGCGCGCTTA 7534
QY 1362 CGACGCGTACAGGCGGACAACTCGATCTCGCGGAGCTGCGGCTCGCGCGCGACGCTG 1421
Db 7535 CGCGCGCGGAGGTGCGAGATCTCTGACAACTATACACGACAACTTCCAGAACGGTTC 7594
QY 1422 CGGTGGTACACCCCTCTGAGCACCCCGCTGGCAGCGCTGAGCAACTAGCTACCGCCAC 1481
Db 7595 CTACGCGGAATCCTCTCTGACGTCTACGACGATACCGCGCGGGGTGTCGGGCAATTTCTA 7654

QY 1806 CACCGTGTCTCCACGACAGTCCACTTCGCCGAGGGCGACACGGGAATCTCGCTC----- 1860
 Db 1789 CACCGTCCACACACAGCCCGCGCTCTTCGCCGTGGAGGTGCGCCCTCCACCGCTCGTCGA 1848
 QY 1861 -TACACCGACGGCGCCCGGCACACTTACCGGCATCGTCGCCGAGATTGGCCAGGC 1919
 Db 1849 GTCCCTGGGGGCTCACGCCCGACCTGCTCGCCGCCACTCCGTGGCGAGATCAGCGCCG 1908
 QY 1920 GATCTAGGGGATGCACACACCGCTCACCGAAGCGCCCGGGGAGACGACGCGCGGA 1979
 Db 1909 CCACGTGCGCGGGTCTGTGCTGCGGACGCCGCCCGCTCGTCGGCGCGCGCGCG 1968
 QY 1980 CAATCGACAGTCTCTGTCG 1999
 Db 1969 CCTCATGACAGGCGTCCCG 1988

Search completed: May 11, 2003, 01:58:54
 Job time : 28690 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 08:41:22 ; Search time 243 Seconds
(without alignments)
10233.138 Million cell updates/sec

Title: US-09-868-328B-3

Perfect score: 2001

Sequence: 1 gcgggtcaccccgacttccc.....atcgacacgtcttcgtgtt 2001

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published_Applications_NA:*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
 - 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
 - 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	97.6	4.9	2712	10	US-09-748-033-4
2	88.2	4.4	882	10	US-09-974-300-735
3	80.4	4.0	15872	9	US-09-860-846-1
4	80.4	4.0	15872	9	US-09-988-384B-1
5	80.4	4.0	15872	10	US-09-861-289-1
6	80	4.0	1896	9	US-10-124-880-15
7	79.2	4.0	2334	10	US-09-476-242-7
8	78.2	3.9	11220	9	US-09-860-846-32
9	78.2	3.9	11220	9	US-09-988-384B-32
10	78.2	3.9	11220	10	US-09-861-289-32
11	78.2	3.9	36778	9	US-09-860-846-5
12	78.2	3.9	36778	10	US-09-861-289-5
13	77.6	3.9	12441	9	US-09-988-384B-3
c 14	77.6	3.9	13613	9	US-09-860-846-3
c 15	77.6	3.9	13613	10	US-09-861-289-3
c 16	77.6	3.9	13613	10	US-09-861-289-3
c 17	75.8	3.8	2322	10	US-09-476-242-18
18	75.8	3.8	2322	10	US-09-476-242-19
19	75.6	3.8	4689	9	US-09-860-846-34

20	75.6	3.8	4689	9	US-09-988-384B-34	Sequence 34, Appl
21	75.6	3.8	4689	10	US-09-861-289-34	Sequence 34, Appl
22	75.6	3.8	13842	9	US-09-860-846-30	Sequence 30, Appl
23	75.6	3.8	13842	9	US-09-988-384B-30	Sequence 30, Appl
24	75.6	3.8	13842	10	US-09-861-289-30	Sequence 30, Appl
25	75.2	3.8	88421	9	US-09-976-059-1	Sequence 1, Appl
26	74.2	3.7	2322	10	US-09-476-242-20	Sequence 20, Appl
27	73.2	3.7	1248	9	US-09-860-846-7	Sequence 7, Appl
28	73.2	3.7	1248	9	US-09-988-384B-7	Sequence 7, Appl
29	73.2	3.7	1248	10	US-09-861-289-7	Sequence 7, Appl
c 30	72.8	3.6	4826	10	US-09-772-304A-1	Sequence 1, Appl
31	69.4	3.5	1294	10	US-09-748-033-2	Sequence 2, Appl
32	65.8	3.3	1458	9	US-09-860-846-9	Sequence 9, Appl
33	65.8	3.3	1458	9	US-09-988-384B-9	Sequence 9, Appl
34	65.8	3.3	1458	10	US-09-861-289-9	Sequence 9, Appl
35	65	3.2	4257	9	US-09-825-288A-1	Sequence 1, Appl
36	64.6	3.2	1107	10	US-09-748-033-6	Sequence 6, Appl
37	64.6	3.2	1140	9	US-09-860-846-15	Sequence 15, Appl
38	64.6	3.2	1140	9	US-09-988-384B-15	Sequence 15, Appl
39	64.6	3.2	1140	10	US-09-861-289-15	Sequence 15, Appl
40	64.4	3.2	1266	9	US-10-145-415-104	Sequence 104, Appl
41	64.4	3.2	3468	9	US-09-988-462-2	Sequence 2, Appl
42	64	3.2	2328	10	US-09-476-242-6	Sequence 6, Appl
43	63.2	3.2	2010	12	US-10-032-717-9	Sequence 9, Appl
44	63.2	3.2	2322	10	US-09-476-242-5	Sequence 5, Appl
45	62.6	3.1	390	10	US-09-790-399-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1

- US-09-748-033-4
- Sequence 4, Application US/09748033
- Patent No. US20020069431A1
- GENERAL INFORMATION:
- APPLICANT: Broadway, Roxanne M.
- APPLICANT: Gongora, Carmenza E.
- TITLE OF INVENTION: EFFECT OF ENDOCHITINASE AND CHITOBIOIDASE AND THEIR
- TITLE OF INVENTION: ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT
- FILE REFERENCE: 19603/3091
- CURRENT APPLICATION NUMBER: US/09/748,033
- CURRENT FILING DATE: 2000-12-22
- PRIOR APPLICATION NUMBER: 60/172,003
- PRIOR FILING DATE: 1999-12-23
- NUMBER OF SEQ ID NOS: 8
- SOFTWARE: PatentIn Ver. 2.1
- SEQ ID NO 4
- LENGTH: 2712
- TYPE: DNA
- ORGANISM: Streptomyces albidoflavus
- US-09-748-033-4

Query Match	4.9%	Score 97.6;	DB 10;	Length 2712;
Best Local Similarity	43.9%	Pred. No. 1.6e-13;		
Matches 650;	Conservative 0;	Mismatches 794;	Indels 36;	Gaps 4;
Qy 466	CGGGCTCGTCCGGTCCGTCTACACATGACGCCGCCAGCGGTGCTCTGGACCCC	525		
Db 845	CGGGCGCGTGGCGCGCTCGTGGCTCGGGTCCCTTCGGCGGGATGTCGGCTC	904		
Qy 526	CAACGCCGGTCCACACCCAGCGGCTTACCAGTGTACTACCTGCACCTCCGACAGAAC	585		
Db 905	CGCGCCCCACCCAGCGCGCGAGCGCGCCGCGCCCTCTACACAGGACG	964		
Qy 586	AACGGCCCCGGGTGGGACACGAGACGAGCGGCGGTGCGCTTCACGACACAC	645		
Db 965	CAGGACTGGGCGCGGTTCGAGGGCAAGTGAGGAAACACACCGACCCGCC	1024		
Qy 646	GGCACGCGTATCGCGTCCGCGCGACTTCGCCGTGTCGGGTGCGGCTCGTCGC	705		
Db 1025	CTACGCGCTGAGCCCTGGAGTGGGACTTCCCGCCCGGAGACAGGTGACCTCGCCTG	1084		

QY 706 ACCGGGAACACGGCAGGTTTCGGCGCGCGCGGTGCTGCGCGCTCGCGACACCGCGAC 765
Db 1085 GACGGCGAGCTCACCACAAACGGCGACCACTGGACCGCTCAAGAACAGAGCTGGCGGG 1144
QY 766 GACGGGTCGCAAGTACACAGGAGTACCTCTACTGCTGACCGACGGCGGTTTACG 825
Db 1145 AGCCTTGGCGCGCGCGCTTCGGTCACTGCTGCTTCAACGGGACCGCGCGCGCACCGCC 1204
QY 826 TTACAGCGCGCTTCGGCGACCGCTTCATCGTCAACACCGAGGTCGGCGCGCACCGCC 885
Db 1205 TCGGGCTGCAAGCTCAACGGCGCTTCCTCGGAGCGGGGAGCGTCCCGCGGACACCGCG 1264
QY 886 GCCGAGATCGAAGACCGCGAGTGGTTCCGCGACCGCCCAAGATCCACTGGGACACCGCGCG 945
Db 1265 CCCACGCGCGCGCGACCG 1324
QY 946 GAGAGATGGTCTGGCTCATCGGACGACTCGGTACGCGCGCTTCTACACCTCGCGGAC 1005
Db 1325 TGAAGCG 1384
QY 1006 CTGCGGCGACTGGACACTTCGCGCGCACTTCGACTACCGCAACCGCGCGCGCGCGCGCG 1062
Db 1385 AAGTGCACCG 1444
QY 1063 -----ATGAGTGGCG 1110
Db 1445 TACTCTCTACTCGGTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1504
QY 1111 CACTGGGTCTCG 1167
Db 1505 GTCAAGGTCAACCG 1564
QY 1168 GCGTACTGGACAGGCG 1227
Db 1565 AACCTCGGCTACTTCAACGAGTGGCGCGTCTACGGCGCGCACTACCGCTCAAAACCTG 1624
QY 1228 TGGCTGAGCTGGGCTGGGACTGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1287
Db 1625 GTGAGCTCGGCTCG 1684
QY 1288 GAGACCAAGCGCTCG 1347
Db 1685 GCGCGAAGTGACCATCGGTGACAGCTTCGCGCGCTTACGCAAGCGGTACACCGCGCG 1744
QY 1348 CCCACGAGCGCTACGAGCGGTACAAACGGCGAGAACTCGATCGTCCGCGAGCTGGCGT 1407
Db 1745 GAGTGGTGGCGCGCTCG 1804
QY 1408 GCGCGACAGCTGGCGCGCTGGTACACCTCTCTGAGACCGCGCGCGCGCGCGCGCGCG 1467
Db 1805 CTCCGCAAGCTCAAGGCCAAGTACCGCGACATCAAGGTCTCTGCTTTCGCGCGCGT 1864
QY 1468 TACGTCACCG 1525
Db 1865 ACTGCTCGCGCGCTTACCG 1924
QY 1526 CATGGAACG 1569
Db 1925 CACGACCTGGTGGAGCG 1984
QY 1570 GACACCG 1629
Db 1985 GAGTACCG 2044
QY 1630 AACATCGGCAAGTACGGAGCAGACCTGTACGTGACCGAGGACCTTCGACCGCTCGCGCG 1689
Db 2045 ATGCTCCAGCGATGCG 2104
QY 1690 TACTCGCTGCG 1749
Db 2105 GACGCGAGCTCG 2164
QY 1750 CACTTGGCGCATCTCTGTCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1809

Db 2165 GACTGGTACAACGTGATGACGTACGACTTCTTGGCGCGCTGGGACAAGACCGCGCGAC 2224
QY 1810 GTGCTCTCCACGAGTCCACTTCCGCGAGGCGGACACGCGGAATCTCGCTCTACACCGAC 1869
Db 2225 GCG 2284
QY 1870 GCG 1909
Db 2285 GCGCGCATCGCGCAAGCTCAAGCGGAGGCGGTCCCGCGCA 2324
RESULT 2
US-09-974-300-735
; Sequence 735, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR APPLICATION NUMBER: 2001-10-05
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 735
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-735
Query Match 4.4%; Score 88.2; DB 10; Length 882;
Best Local Similarity 63.4%; Pred. No. 2.4e-11;
Matches 135; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
QY 1055 TCGCGCGCATCGAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1114
Db 654 TCGGATCATCGAATGTCGCGATCTGTATAGATGGGCGGATGACGGAACATATAAT 713
QY 1115 GGGTGTCTCG 1174
Db 714 GGGTCTTGGAGCAAGCGCGCAACGGGAAAGGCGCGCGCGCGCGCGCGCGCGCGCG 773
QY 1175 GGACAGGCG 1234
Db 774 GGACAGGCG 833
QY 1235 ACTGGGCTGGGACTGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1267
Db 834 ATTACGGAATTGACTGGTATGCGGACGTGACAT 866
RESULT 3
US-09-860-846-1
; Sequence 1, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-1

Query Match      4.0%; Score 80.4; DB 9; Length 15872;
Best Local Similarity 42.7%; Pred. No. 1e-09;
Matches 649; Conservative 0; Mismatches 856; Indels 15; Gaps 4;

QY 492 CATGACGCCCCCAGCGGTGCTCTGACACCCCAAGCGCCGCTACACACCGGCGC 551
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 472 CATGCCAACCCCGCTCTGTACACCTCGGCTGAGGGCCGAGCCTCACCGTGCAGC 531
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 552 CTACACGCTGTACTACCTGCACTCGACAGCAACAGCGCCCGGGTGGAGCACCGC 611
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 532 CGCGAGTGTCTCGCTGCTGCGCGTGCACCTGCGCTGCGAGTCTCTGCGCGCGGGA 591
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 612 GAGCAGACCGGCGTGGCTTCACGACACACGCGCACCGCTGATGCCGCTGCGGCCGA 671
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 592 GTCCACGAGCGGCTGCTGCGCGGCTGAACCTCAACATCTCTCGCGGAGCGCGTGAC 651
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 672 CTTCCCGTGTGCTCGGCTGCGGCTGCTGCGGACCGGGAACACGAGGTTTCGGCGC 731
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 652 GGAGGAGGCTTCGGTGGACTCTCCCGGAGCGCACCGCTACACCTTCGAGCGCGGCG 711
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 732 CGGCGCGTGTGCGCGCTCGGACACGCGGACCGGCTCCGCAAGTACCGAGGCA 791
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 712 CACGGATGTCTCGGCGGAGGGCGGAGTGTCTGCTACTCAAGCGGCTCTCCGCGCG 771
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 792 GTACCTTACTGTGCGACGCGGCGGTTCACGTTTACCGGCTTCGCGGACCGCGTAT 851
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 772 CCTCGCGGACGCGACCGTGTCCACGCGTCTATCGCGGCGCGCTCAACACGACGG 831
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 852 CGTAAACACGAGTGTGCGCGCGCCACACGCGCCCGC---GAGATCGAGAACGCGGAGTG 908
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 832 AGCCACCCCGGCTCTACCGTGTCCAGGAGGCGCCCGGAGAGAGGCTCTCGCGGAGCG 891
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 909 GTTCGCGGACCCCAAGATCTCACTGGGACACGCGCGCGGAGATGGTCTGCTGATCGG 968
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 892 GTACCGGAGGCGGCGCTTGACCGCTGCGCGTCCAGTACGTGCACTCCACGCGACCGG 951
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 969 AGGACTGCGGTACGCGCGGTTCATACCTTCGCGGACCTTCGCGGACTGCGGACTTCGCGG 1028
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 952 AACCCCGCTGCGCGACCCCATCAGAGCGCGCGCTCGCGCGCGTCTCTCGGCTCGCGGCGG 1011
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1029 CAACCTTCGACTACCGAACCGCCCTCGGCGGATCGAGTCCCGGACCTGTTGAGAT 1088
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1012 CCGCGCGGACGAACCCCTGCTGCTGCGGTGCGGCAAGACGAACTCGGCGACCTCGAAGG 1071
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1089 CACCGCAG---ACGACGGGACACGCACTGGGTGCTGCGCGGACGATGGAGCGCTACGG 1145
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1072 CGCGCGCGGCTGCTGCGGCTCATACAGCGTCTCTCGGCTGCGCGGCGCGGATCC 1131
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1146 CATCGGCTCTCCCATGAGCTAGCGCTTACGAGGACCTTGGGACGCGGAGCACTTCCA 1205
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1132 GCGGAGCTCACTTCCGTACGCGCCACCGGACATCCGCTCGACACCTCTCGGCTCGA 1191
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1206 CGCGGACGACTCACCGCGCAATGCTCGATGGGCTGAGGACTGTTAGCGGCGCGTAC 1265
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1192 CGTGGCGGACGCGCTGCGGAGTGGCGGACCGCGGACCGGAACTCTCTCGCGGCGGTGAG 1251
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1266 CTGGCATCGATCGAGCGCGGAGACCAAGCGCTCGCATCGCTGATGAACAATG 1325
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1252 CTCGTTGCGATGCGGCGGACCAACGCGCTCTCTGAGGAGGCGCGCGCGGAGG 1311
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1326 GAAGTACCGCGGACGAGCTGCGGACCGGACGATCCGAGCGGCTACACGGGAGCAATC 1385
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1312 CGCGGAGGACCGCGCATGATGAGGAGACCCCGCTCGACAGCGGCGCGACTGCCCTT 1371
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1386 GATCGTCCGAGTGTGCGGCTGCGCGGACAGCGCTGCGGCTGGTGTACACCTCTCTGAGCAC 1445
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 4

```
US-09-988-384B-1
; Sequence 1, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988.384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-1
```

```
Query Match      4.0%; Score 80.4; DB 9; Length 15872;
Best Local Similarity 42.7%; Pred. No. 1e-09;
Matches 649; Conservative 0; Mismatches 856; Indels 15; Gaps 4;

QY 492 CATGACGCCCCCAGCGGTGCTCTGACACCCCAAGCGCCGCTACACACCGGCGC 551
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 472 CATGCCAACCGGCTCTCGTACCACCTCGGCTGCGAGGCGCCGAGCCTCACCGTCGACGC 531
QY 552 CTACAGCTGTACTACTCTGACTCCGACAGAACAGGCGCCGCGGTGGGACACAGC 611
Db 532 CGGCGAGTGTCTGCTGCTGCGCGTGCACCTGGCGTCCGAGTCCCTGCGCGCCGGGA 591
QY 612 GAGCAGACCGGCGGTGCGCTTACGACACACGACGACCGTGTATGCGCGTGGGCCGA 671
Db 592 GTCCACAGCGGCGCTGCTGCGCGGCTGAACTCAACATCTCGCGGAGAGCGCGGTGAC 651
QY 672 CTTCCCGGTGTGTGCTGCGGCTGCGGACCGCGAAGACGAGGTTTCGGCGC 731
Db 652 GGAGAGCGCTTCTCGGTGACTCTCCCGGAGCGACCGCTACACCTTTCGACGCGCGGC 711
QY 732 CGGCGCGGTGCTGCGCTCGCGACCGACGACGCGCGGTGCGGACGCGTCCGACGAGGA 791
Db 712 CAAGGATTCGTCGCGGCGAGGCGGCGGAGTGTGCTGCTACTCAAGCGCTCTCCGCGC 771
QY 792 GTACCTCTACTGCTGACCGAGCGGGGTTACGTTTACCGCGCTCGCCGACCGCGTAT 851
Db 772 CTTGCGCGAGCGACCGGTGTCTCAGCGGCTCATCGCGGCTATCGCGGCGACGCGTCAACAGACGG 831
QY 852 CGTCAACACCGAGTGTGCGCGCGCCACACGCGCGC---GAGATCGAGAACCGCGAGTG 908
Db 832 AGCCACCGCGGTCTACCGTGCCTCCAGAGCGCGCGCCAGGAGAGGTGCTGCGGAGGC 891
QY 909 GTTCGCGACCCCAAGATCACTGGGACACCGCGCGGAGAAATGGGTCTGCGTCACTCG 968
Db 892 GTACCGAAGCGGCGCTGACCGCGTCCGCGCTCCAGTACGTGCAACTCCACGCGACCGG 951
QY 969 AGGACTGCGGTAGCGCGGCTTCTACACTCGCGCGGAACCTGCGGAGTGGACACTTCGCGC 1028
Db 952 AACCCTCGTGGCGAGACCCCATCGAGGCGCGCGCTGCGCGCGCTCTCGGCTCGCGCGC 1011
QY 1029 CAATCTCGACTACCGGAACCGCTCGCGCGCTGAGTGTGCGCGACCTGTGCGAGAT 1088
Db 1012 CCGCGGAGAGACCCCTCTCTGCTGCGTGGCGGCGGAGAGAGTGTGCGGACCTCGAAG 1071
QY 1089 CACCGGAG---ACGAGGAGACAGCGACTGGGTGTCTGCGCGGACGATGAGCGCTACGG 1145
Db 1072 CGCGCGCGCATCGTGGCTCATCAAGACGCTCTCGCGCTCGCGCGCGCGGATCCC 1131
QY 1146 CATCGGCTCCCATGACGTACGCTTACTGGACAGGACCTGGGAGCGGCGAGCAGTTCA 1205
Db 1132 GCGGAGCTCAACTTCCGTACGCGCGCACCGGACATCCCGCTCGACACCTCGGCGTCA 1191
QY 1206 CGCGACGACCTCACCGCGCAATGTCTGACTGGGCTGGGACTGTGACGCGCGCTCAC 1265
Db 1192 CGTGGCGAGCGCTCGGAGTGGCGGACCGCGACCGGACCGGACTCTCGCGCGGTCA 1251
QY 1266 CTGGCGCATGATGACGCGCGCGAGACCAAGCGCTCGGCTCGGATGATGAACAATG 1325
Db 1252 CTCGTTGCGCATGGGCGGCGACCAAGCGCGCATCGTCTCTCAAGGAGGCGCGCGCAGG 1311
QY 1326 GAAGTACCGCGACGGAGCTCCCAACCGACCATCCGAGCGCTACAGCGGCGAATC 1385
Db 1312 CGGCGAGACGCGCGCATGATGAGGAGACCGCGCTCGACAGCGGCGCGCTCCCTT 1371
QY 1386 GATCGTCCGAGCTGCGGCTGCGCGACAGCTGCGGCTGGTACACCTCTCTGAGAC 1445
Db 1372 CTTGTCACCGCGCGGCGGAGGCGCTGCG---GCGCGGAGCGCGCGCTCGACA 1428
QY 1446 CCGCTGGAGCGGTGAGAACTAGCTACCGCGGACCGGCTCGGCTCGGATGAGACCGGT 1505
Db 1429 GCGCGTCAAGGAGGACCGGAGCTGCGCGCGCGGCGGCTCGGCTGCTGCTGCTGCTG 1488
QY 1506 CGACGCGAGCGGCTCTGCTCCATGAAAGGAGCGCGCATACGAGATCGAGCTGACATCC 1565
Db 1489 CCGTAGGTTTACGACACCGGTGCGGTGCTCTCGCGCGGACCGCGCGCGCTCTCGA 1548
QY 1566 GTGGGACACCGGAGGAGCTGCGGATCTCGGTGCGCGCTCCCGCGGAGGAGGAGGAGG 1625
Db 1549 CGGCGCTGCGGCGCTCGCGCGGAGCGCGCGCGCGCGGCTGCTACCGGACCGCGCGC 1608
```

RESULT 5

```
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US2002110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1
```

Query Match 4.0%; Score 80.4; DB 10; Length 15872;

Best Local Similarity 42.7%; Pred. No. 1e-09; Mismatches 856; Indels 15; Gaps 4;

Matches 649; Conservative 0;

```
QY 492 CATGACGCGCGCGCGGCTGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 551
Db 472 CATGCCAACCGGCTCTCGTACCACCTCGGCGCTGCGAGGCGCGGCTCACCGTCGACGC 531
QY 552 CTACAGCTGTACTACTCTGACTCCGACAGAACAGGCGCCGCGGTGGGACACAGC 611
Db 532 CGGCGAGTGTCTGCTGCTGCGCGTGCACCTGGCGTGGCGTCCCTGCGCGCGCGGGA 591
QY 612 GAGCAGACCGGCGGTGCGCTTACGACACACGACGCGCGCGCGCGCGCGCGCGCGCG 671
Db 592 GTCCACAGCGGCGCTGCTGCGCGGCTGAACTCAACATCTCGCGGAGAGCGCGGTGAC 651
QY 672 CTTCCCGGTGTGCTGCGGCTGCGGCTGCGTCCCGCGGAGGAGGAGGAGGAGGAGG 731
Db 652 GGAGGAGCGCTTCTGCTGAGACTCTCCCGGAGCGCGCTACACCTTTCGACGCGCGGC 711
```



```
QY 732 CGGCGGGTCTGTCGGCTCGCGACCCAGCGGACCGGCGTCCGCAAGTACCAAGAGCA 791
DB 712 CAACGATTTCGTCGGGGGAGGGCGGGAGTCTGCTACTCAAGCGGCTCTCCGCGC 771
QY 792 GTACTCTACTGTCGACCGGCGGGTTCAGGTTTACCGCTTCCCGGACCCCGTCTAT 851
DB 772 CTTGCCGACGGGACCGTGTCTCAGCGGCTATCCGCGCCAGCGGCTTCAACACGAGCG 831
QY 852 GTTCAACACCGACGTCGCGCCGACCCAGCGGCGGCTTCCGAGTTCGAGTTCGAGT 908
DB 832 AGCCACCCGGGTCTCACCGTGTCCAGCAGCAGGCGCGCCGAGAGAGTGTCTCCGCGAGC 891
QY 909 GTTTCGCGACCCCAAGATTCACCTGGGACACCGCGCGCGGAGAAATGGTCTCATCGG 968
DB 892 GTACCGGAAGCGGCGCTTGGACCGTCCGCGGCTCCAGTACGTCGAACTCCACGGCACCG 951
QY 969 ACGACTGCGGTACGCGCGTTCACACTCGCGGACCTCGCGGACCTGGACACTTCGCGG 1028
DB 952 AACCCCGTTCGGGACCCATCGAGGCGCGCGCTCGGCGCGTCTCGGCTCGGCGCG 1011
QY 1029 CAACTTTCGACTACCCGGAACCAACCGCTCGGGGCTAGTTCGCGGACCTGTTCGAGAT 1088
DB 1012 CCCCGGACGACCCCTGCTCTGCGGCTCGGCCAAGACGAGTTCGGGACCTCGAAGG 1071
QY 1089 CACCGCAG---ACGACGGGACACGCCACTGGGTGCTCGCGCCAGCATGGACGCTACGG 1145
DB 1072 CGCGCGCGGCTCGTTCGGGCTCATCAAGACGCTCTCGCGCTCGGCGCGGCGGATCCC 1131
QY 1146 CATCGGCTCCCAATGACGTACGCTACTGGACAGGACCTGGGACGGGAGCACTTCCA 1205
DB 1132 GCGGAGCTCAACTTCCGTACGCGCCACCGGACATCCGCTCGACACCCCTCGGCTCGA 1191
QY 1206 GCGGACGACCTCAACCCGCAATGCTGACTGGGCTTGGGACTGGTACCGGCGGCTCAC 1265
DB 1192 CGTGGCCGACGCGCTGCGGGAGTGGCGCACCGGACCGGAACTCTCTCGCGGGGTGAC 1251
QY 1266 CTGGCCATCGATCAGCGCGCGGAGACCAAGGCTCGCCATCGCGTGGATGAACAAGT 1325
DB 1252 CTGCTTCGGCATGGCGCGCACCAAGCGGCTCTCTCGCGGAAAGCGCGGCGGCGG 1311
QY 1326 GAAGTACGCGCACGCGACGTCCCAACGACGATCCGACGCTACACGGGCAACATC 1385
DB 1312 CGGCGAGCAGCGGCTATGATGAGGAGACCCGCTCGACAGCGGGCGGCTACTGCCCTT 1371
QY 1386 GATCGTCCGCGAGTTCGGGCTCGCGGACAGCTTGGGCGCTGGTACACCTCTCGAGCAC 1445
DB 1372 CGTGTACCGCGCGCGGCGGAGGCGCTGCG---GCGCCGACGGCGCGGCGCTTCACGA 1428
QY 1446 CCCGTGGCAGGCTGACGACACTAGTCAACGCGCACACGACACTTCCCGGACCGGACCGT 1505
DB 1429 GCGCGTCAAGCGGACCGGAGCTCGCGCGCGCGGCACTCGCGCGGCTGGTTCACCA 1488
QY 1506 CGAGGACGCGCGTTCGCTGCAATGGAACGAGCGGCGCATACAGATTCGAGTTCGACATCGC 1565
DB 1489 CCGTACGGTCTTCAGCAGCCGCTGCTGCTCGCGCGGACCGCGCGGCTCTCTCGA 1548
QY 1566 CTGGGACACCGGACGAGTTCGGGCTCTGCTGGGCGGCTTCCCGGCAACCGGCA 1625
DB 1549 CGGCTTCGCGCGCTCGCGGCGGAGCGCGCGCGCGGCTGGTTCACGGGACCGCGCGC 1608
QY 1626 CACGAACATCGGCAAGTACGAGACGACTGTACGTGACGAGGAGACCTCCGACCTCGC 1685
DB 1609 CCGCGGCGGCTCGCGGCTCTGTTTCAGCGGCGAGGTTGCCAAGCTACGGGCTATGGGAT 1668
QY 1686 CGGCTACTCGTCCGCTTACTCGGAGCGCGCGCGGCTTCCCGGACCGCGGCGCGGCT 1745
DB 1669 GGAGTGTACGCGCGGCGGCGGCTTTCGCGACGCGCTTCGACGCGCTCGCGCGGCGGACT 1728
QY 1746 GGTGACCTGCGCATCTCTGTCGACACCGAGCGCTTCGAGGCTTCTGTCACCGCGGCA 1805
DB 1729 GGACCCCTCTCTGACCGCGGCGGCTCGCGGAGTCTCGCGGGGCGGACACCTCGACCG 1788
QY 1806 CACCGTGTCTCCACGAGGTCCACTTCGCGGAGGCGGACACGCGGAATCTCGCTC---- 1860
```

```
DB 1789 CACCGTCCACACACAGCCGCGCTCTTCCGCTGGAGGTGCGCCCTCCACCGCTCTCGA 1848
QY 1861 -TACACCGAGCGCGCGCGACACTTCACCGGCTATCGTCCGCGAGATTGGCCAGCG 1919
DB 1849 GTCCCTGGGCGTCAAGCGCGGACCTGCTCGCGGCGCACTCCGTCGGGAGATCAGCGCGC 1908
QY 1920 GATCTAGGCGATGACACACACCGCTCACCGAAAGCGCGCGCGGAGAGACGACGCGCA 1979
DB 1909 CCACGTCCGCGGGTCTTGTGCTGCGGAGCGCGCGGCTCTGTCGCGGCGCGCGCG 1968
QY 1980 CAATCGACACGCTCTGTCG 1999
DB 1969 CCTCATGAGCGCTCCCG 1988

RESULT 6
US-10-124-880-15
; Sequence 15, Application US/10124880
; Publication No. US20030026810A1
; GENERAL INFORMATION:
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene No. US20030026810A1boe
; APPLICANT: Schulein, Martin
; APPLICANT: Outtrup, Helle
; TITLE OF INVENTION: No. US20030026810A1el Rhamnogalacturonan Hydrolases
; FILE REFERENCE: 5572.204-US
; CURRENT APPLICATION NUMBER: US/10/124,880
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US/09/311,626B
; PRIOR FILING DATE: 1999-05-13
; PRIOR APPLICATION NUMBER: 0608/98
; PRIOR FILING DATE: 1998-05-01
; PRIOR APPLICATION NUMBER: 60/084,358
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1896
; TYPE: DNA
; ORGANISM: Streptomyces coelicolor
US-10-124-880-15

Query Match 4.0%; Score 80; DB 9; Length 1896;
Best Local Similarity 43.8%; Pred. No. 1.6e-09;
Matches 665; Conservative 0; Mismatches 825; Indels 30; Gaps 6;

QY 489 CCACATGACGCGCGCGGCTGCTGCGACCCCGGCGGTCACCGCGGTCACCGACCGG 548
DB 276 CAACTCCAGCGCGCTACCGGCTCCACCACTTCTTCCACTCCGCGCGCGCTCCACGC 335
QY 549 CGCCTACCACTGCTACTACTTCCACTCCGACAGAACACCGCGCGCGGTCGAGTCCG 608
DB 336 CGACTACACCGTCCGCGGCTGTAAGCGGACGAGGCGGCTCGTCCGTCACCGCAT 395
QY 609 CGCAGACGACGACGCGGCTGCTTACGACCAACACCGCGCGCGGTCGAGTCCGCTCG 668
DB 396 CGAGTTCGCGCGGCTACAGGACGATACCGATACGCGCGCGCTCCGCGGCGGACCG 455
QY 669 CGACTTCCCGGCTGCTGCGGCTCGCGGTCGCGGACCGGAGAACACGCGGAGGTTGCG 728
DB 456 CGACGCGGCTCTCTACACTACGAGGCGCAACGACGCGCTCCGTCGCGGACCTCGAG 515
QY 729 CGCGCGGCGGTCGTCGCGCTCGGACCGCGGCGGTCGCGGCGGTCGCGCAAGTACCA 788
DB 516 CGCGCGCGCTCGACCTCGCTCAAGTGGACGCGCGGCAACGCGGAGCAACTCCAGTC 575
QY 789 GCAGTACCTCTACTGTCGACCGAGCGGCTTACGTTTACCGCTTCCCGGACCGCGCT 848
DB 576 CGGCTACACCGGCAACGCGTCTGACGGGATCAAGCTCGACGCGA---CCCGCTGTG 632
QY 849 CATGCTCAACCGGAGTTCGCGCGCGGCGGCGGACCGCGCGGAGATCGAGACCGCGAGTG 908
```

Db 633 GCGCGTCCGACCTGGCGCGCAACATCCGCTCCGCGCCCACTACACCCAGTTCCAGGTGA 592
QY 909 GTTCCCGGACCCCAAGATCACTGGGACACCGCGCGGAGATGGTCTGGGTCAATCGG 968
Db 693 CGACTACGAGCGGCGCGCGCGGCGAGGTGCGCATGAGACCGCGCGACGCCAACGA 752
QY 969 AGCACTCGCGTACGCGCGCTTCTACACCTCGCGGAACCTGCGGACTGGACACTTCGCGG 1028
Db 753 CGGACCGCGCGGTATCGGCAACTCTCTCGCGGATCACCGCACTCGAGCGGTACGT 812
QY 1029 CAATTCGACTACCCGAACACCGCCTCGGGGCGATCGAGTGGCCGACCTGTTCCAGAT 1088
Db 813 CTTCTCGCGCGCGAATACCTCACCATGTTCAACGCGCGGACCGGACCGATGGGAC 872
QY 1089 CACCGAGACGAGCGGACACGCACTGGTGTGCGC---GCCAGATGAGCCCTACGG 1145
Db 873 CTGCGACTAGTCCGCGCGCGGCTCGGTCTCTCTCGGGGACTCTCTAGCGCAACCG 932
QY 1146 CATCGCGCTCCCATGACGTACGCTTACTGGACAGGACCTGGGACGGGAGAGTTCCA 1205
Db 933 CFTCGACCGGTTCTTGGCGGACGCGGTACCTGGACGCTCCCGCCCTCGGTGATTAT 992
QY 1206 CCGGACGACCTCACCCCGCAATGGCTCGACTGGGCTGGGACTGTACGCGCGCTAC 1265
Db 993 GCGCGCGGTACTACACCGGACGGTGTATCGCGGCTGGGACTG---CGGACCGCGG 1049
QY 1266 CTGGCCATCGATCGAGCGCGCGGAGACCAAGCGCCTCGGCATCGCTGGATGAACAATG 1325
Db 1050 GTTACCGCGCGCTGGACCTTCGACACCAACTCTCTCCACCAACGCGGCAAGGCTACA 1109
QY 1326 GAAGTACCGCGGACGGAGTCCCAACGACGATCCGACGCTTACAGCGGAGAACTC 1385
Db 1110 CGGCCAGGGCAACCACTCTCGCTCGCGGACGTGGACGTGACGGCGCGGAGAGAT 1169
QY 1386 GATCGTCCGAGCTGGGCTCGCGGACAGCTGGCGCTGTACACCTCTGAGCAC 1445
Db 1170 CTTCTACGCGCGGATGGCGCTCGAGACAAACGGCTACGCGCTGTGGACCAACGAGAACG 1229
QY 1446 CCGCGTGGCAGCGCTGACAACTAGTCAACCGCCACACAC---ACACTCCCGGACCG 1499
Db 1230 CCACGGGACGCCATGCACGTGGCGACCTCGACCGTCCCGGCGGCGCTGGAGGAT 1289
QY 1500 GACCGTACGCGGAGCGGCTCTCGCATGGAACGAGACCGCATACGAGATCAGCTCA 1559
Db 1290 CAAGTTCGACGAGGACGGCTGGAAGCCCTCGTCACTTGGGCGACCGCGGACGGCA 1349
QY 1560 CATCGCTGGGACACGCGGACAACTCGGCATCTCGGTGGCGCGCTCCCGGACGGAAC 1619
Db 1350 GATCCTCTGTCCACCGGCGGAGCGGCGACAAAGCGCGGTCTCGGGGACATCTG 1409
QY 1620 CCGGACACGAAATCGGCAAGTACGGAGACAGCTGTACGTGACCGGAGACCTCCGA 1679
Db 1410 GTCGGCGAGCGGCGCGGAGTCTGTGTC-----GTCCGCGGAGCGGCGATCCGCA 1463
QY 1680 CTTCCCGGCTACTCGCTCGCCCTTACTCGGAGCGCGCGCCCACTACACCGCGCGC 1739
Db 1464 CCCCAGGCGACCGTGTGCGGAGCGGCAAGCCCTCCAGCGCAACTCTCTTCTGTG 1523
QY 1740 CGGATCCGTGCACTCGCGCATCTCTGTCACACCGACGAGCGTCTGCTGCAACCG 1799
Db 1524 GGACGGGACACCGTCCGTGAACTCTCGACGGCACCCACGTGCAAGTACGGACCTC 1583
QY 1800 CGGCGACACCGTCTCTCCAGAGTCCACTTCCCGGAGCGGACACGGGAATCTCGCT 1859
Db 1584 GGGCGACA-----CCCGCTGTCTACCGGCTCGCGCTCTCCAAACACGGCAC 1634
QY 1860 CTACACCGAGCGGCGCGGACACTTACCGGATCTGCTCGCGGAGATTGGCAGCG 1919
Db 1635 CAAGGCCACCGGCTGTGGCGGCGGACATCTCTCGGCGACTGGCGGAGGTCTGTG 1694
QY 1920 GATCTAGGCGATGCACACACACCGCTACCGAAGCGCGCGCGCGGAGACGCGCA 1979

Db 1695 GCGCAGCTGAACAACACGCGCCCTGCGCATCTACTCCACCCCTACGACACGACACCG 1754
QY 1980 CAATCGACACGCTCTCGTGC 1999
Db 1755 CATCAGCACCTCTCTCCAG 1774

RESULT 7
US-09-476-242-7
; Sequence 7, Application US/09476242
; Patent No. US20020146683A1
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: MARTIN, Karin
; APPLICANT: HARTOG, Eric
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
; FILE REFERENCE: 1605,002
; CURRENT APPLICATION NUMBER: US/09/476,242
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2334
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Leul22-Ser199
US-09-476-242-7

Query Match 4.0%; Score 79.2; DB 10; Length 2334;
Best Local Similarity 43.8%; Pred. No. 2.4e-093;
Matches 458; Conservative 0; Mismatches 573; Indels 15; Gaps 2;

QY 795 CCTCTACTGCTGACGACGCGGGTTCACGTTACCGCCCTGCGCGACCCCGTCATCGT 854
Db 147 CTGTCTCGCGGACGCGCAAGCCCTACGACACCGAGGTGCACAACGTGTGGGCCAC 206
QY 855 CAACACCGAGGTTCGCGCGCGCCACACGCGCGCGAGATCGAGAACGCGGAGTTCG 914
Db 207 CCACGCGCTCGTGCCACCGACCGCCCAACCCCGAGGAGATCGTGGAGAACGTGACCGA 266
QY 915 CGACCCCAAGATCCACTGGGACACCGCGCGGAGATGGTCTGCTGATCGGACGACT 974
Db 267 GAATTTCAATGTGGAAGAACAACATGTGGAGAGATGACGAGGACATCATAGCT 326
QY 975 GGGGTACGCGGGTTCACACCTCGCGGACCTGCGGACTGGACACTTCGCGCGCACTT 1034
Db 327 GTGGGACACAGCCCTGAAGCCCTGCGTGAAGCTGGGCAACAGCGTATCACCCAGGCTG 386
QY 1035 CGACTACCGGACCGCCCTCGGGGCGATCGAGTCCCGGACCTGTTGAGATCACCG 1094
Db 387 CCCCAGGTGAGCTTGGAGCCCATCCCCATCTACTGCGCCCGCGCGGCTTCGCCA - 445
QY 1095 AGACGACGGACACGCACTGGTGTCTCGCGGACGCGCTAGGACGCTAGGATCGGCT 1154
Db 446 -----TCCTGAAGTGCACGACAAGAGTTCACGCGCGGCGCCCTGCAACACGTGAG 500
QY 1155 CCCCATGAGTACGCTACTTGACAGGCACTTCGGGACGCGGAGAGTTCACCGCGGACGA 1214
Db 501 CACCGTGCAGTGCACCCACGGCATCGCGCGGTGGTGGAGCACCCAGCTGCTGTGAACGG 560
QY 1215 CTTACCCCGCAATGGCTCGACTGGGCTGGGACTGGTACGCGGCGCTCACCTGGGCCATC 1274
Db 561 CAGCTTGGCGGAGGCGGTGTGATCGGACGAGAACTTACCGCAACGCGCCCAAGAC 620
QY 1275 GATGACGCGCGGAGACCAAGCGCTCGCCATCGGCTGGATGAACAACCTGGAAGTACGC 1334
Db 621 CATCATCTGTCAGTGAAGGAGAGCTGGAGATCACTGCACCCCGCCCAACACACAC 680
QY 1335 CGACGCGAGCTCCCGACCGGACGATCCGACGGCTACACGGGCGAGAACTCGATCGTCCG 1394
Db 681 CCGCAAGAGCATCACCATCGGCGCGCGCTTCTACGCCACCGCGGAGACATCATCG 740

Db 4000 GTGACCTCGGTGGAGCCCTGGGAGGCGACCGCGGTCAACGAGGTGCGACCGGG 4059
QY 1780 GTGAGGCTTTCGTAACCGCGCGGCACACCGGTCTCTCCAGAGGTCCACATTCGCGGAG 1839
Db 4060 GAGCGGTGCGCGCTGCGGCTGCGCCCTCGCCCGCGGACGCGCTCAACCGCGCTG 4119
QY 1840 GCGGACCGGGAATCTCGGTCTACACGAGCGCGCGCGCGACATTCACCGG 1892
Db 4120 GACACCGCGCTGCGGACGCGGACACCGCGGTACGATCGCGGAGGTGACTG 4172

RESULT 9

US-09-888-384B-32
; Sequence 32, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-888-384B-32

Query Match 3.9%; Score 78.2; DB 9; Length 11220;
Best Local Similarity 41.9%; Pred. No. 3.4e-09;
Matches 601; Conservative 0; Mismatches 82; Indels 9; Gaps 2;

QY 469 GCGTCTCGTCCGCTTACACATGACGCCCCCAGCGGTGCTCTGGACACCCCAA 528
Db 2740 GACTCCCCCGCTCAGGGCTCCGTGAGGACTCCTGGCGTACCGACTGGAAG 2799
QY 529 CGCCGGTCAACACCGCGCGCTTACCAGTGTACTACCTGCATCTCGACGAGAAC 588
Db 2800 CGCTCGCGGTGCGGACGCGTCCGAGCGCGCGGCTGTCGGGCGTGTCTGCTC 2859
QY 589 GCGCCGCGGCTGGGACACGCGAGCAGCAGCGCGTCCGCTTACGACACCGCGC 648
Db 2860 GTCCCGGAGGACCTTCCGCGGAGCGCGCGGCTGCTGCGCGCTGTCGGCGCGC 2919
QY 649 ACCGTGATGCCGCTGCGCGCGGCTTCCCGGTGTTGTTGCGGTGCGGCTGCGGAC 708
Db 2920 GCGGACCGGTACAGCTGAGGTGTCCTCGGTGGCGAGCGGCGCTGCGCGCGAG 2979
QY 709 GCGGACAGCGC---AGGTTTGGCGCGCGGCGGTGCTGCGGCTGCGGACCGCGG 765
Db 2980 CTGGCGGAGGCGCTGGCGGCGCGGCTGAGCGCTGAGCGGCTGCTGCTGCTGCG 3039
QY 766 GAGCGGTCCGCACTACAGAGCAGTACTCTACTGTCGACCGCGGCGGTTCAGG 825
Db 3040 TGGGACGAGAGCGCGCACCCCGCGCACCCCGCGCTTACCCGCGGCGACCGCGCG 3099
QY 826 TTACCGCGCTGCGCGACCGCGCTATCTCAACACGAGGTGCGCGCGCGCACCGCGC 885
Db 3100 CTCACCTGTTGAGCGCGCTGAGGACCGCGCGGTGCGCGCGCTGTTGTTGCTGAC 3159
QY 886 GCGGAGATCGAGACCGCGGTGTTCCGCGGACCGCGGACGAGTCCACTGGGACACCGCGC 945
Db 3160 CACGCGCGGTGTTGCTGCGGCGGCGCGGCGGACGCTCACTCCCGCGCGCGGCGGTG 3219
QY 946 GGAGAATGGGTCTGCTGTCATCGGAGACTGCGGTACGCGCGGCTTCTACACCTCGCGGAC 1005

Db 3220 TGGGCACTGGGCGCGGTGCGCGCTCGCGCGCTTGGAGGACACCCCGAGCGGTGGGCGCGCTGATCGAC 3279
QY 1006 CTGCGCGACTGGACACTTTCGCGCGCAACTTCGACTACTCCGGAACACCGCCTCTCGCGCGCATC 1065
Db 3280 CTGCGCTCGGAGCGGACCGGCGCGCTTGGACCGCATGACACACGCTCTCGCGCGCGGT 3339
QY 1066 GAGTCCCGCGGCTTTCGAGATCACCGAGACGAGCGGACGACGACCTGCTGCTGCTGCTGCTG 1125
Db 3340 ACGGTGAGGACGAGTCCGCTGCGCTGCGCGCTGCTGCGCGCGCTGCTGCTGCTGCTG 3399
QY 1126 GCCAGCTGAGCGCTACGCGATCGCGCTCCCATGAGTACGCTACTTGGACAGCGAC 1185
Db 3400 GCTTCCCTCCCGCGCGACGCGACGCTTTCGCGCTGTTGCGAGCGCGAGGCGAGGCTGCTC 3459
QY 1186 TGGGACGCGGAGCTTTCACGCGCGACGACCTTACCCCGCAATGGCTCGACTGGGCGTGG 1245
Db 3460 GTCACCGGTGCGGAGGAGCTTGGCGCGCGGAGCGCGCGCTGCGCGCGGACGCGC 3519
QY 1246 GACTGTAGCGCGCGCTACCTTGGCCATCGATGAGCGCGCGGAGCGCGCAAGCGCGCTCGCC 1305
Db 3520 GCGGACACCTCTCTCTCCACACCGCGCTCGCGCAGCGAGGCGCGGAGGCGAGGCGACCTCC 3579
QY 1306 ATCGGTGATGAACAACCTGGAAGTACCGCGACGCGAGCTCCCGCACGACGACGACATCCGAC 1365
Db 3580 GGTGCGCGCGGAGGACTCGCGCTCCCGGCTGCTGCGGAACTCGCGGACCTGGGCGCG 3639
QY 1366 GGCTACAAAGCGGCGAGACTCGATCTCGCGGAGCTGCGCGCTGCGCGCGAGCG-CT 1419
Db 3640 ACGGCGACCGTCTGACCTTGGACCTACGAGCGCGGAGGCGCGCGCGCTGCTGCGC 3699
QY 1420 GCGGCTGTTACACCTCTCTGAGCACCGCGGTGGAGCGGTGAGCACTACGTCACCGCG 1479
Db 3700 GCGGTCTCCGACGCGACCGCTAGCGCGCTCTCTCCACCTGCGCGCGCGCGCTGACCTCC 3759
QY 1480 ACCACACACTCCCGGACCGGCGGTGCGAGCGCGCTCTCTGCTGCGCGCGCGCGCG 1539
Db 3760 GAGCGCTCGCGCGGACCGCGCGGCGCGCTGCGCGCTGCTGCGCGCGGAGCGCGAC 3819
QY 1540 GCATACGAGATCGAGCTCGACATCGCTGGGACACCGCGGAGCGTGGGCTGCTGCGGTG 1599
Db 3820 GCGCGCTTCCACTGAGCGCTCTCTGCGGAGCGCGCGGTGCGGAGCGCGCTGCGCGCG 3879
QY 1600 GCGCGCTCCCGGAGGAAACCGCGCACACGACATCGGCAAGTACGAGGAGCGAGCTGAC 1659
Db 3880 GTCTGTGCTCTCTCTCTGCGTGGCGGCTGCGGCGCGCGGTGCGGCGCGGTGCGGCG 3939
QY 1660 GTGACCGGAGGACCTCGGAGCTGCGCGGTGCTGCGCGCGCGCTTCTGCGGAGCGCG 1719
Db 3940 GCGCGGTACGCGCTTCTTCTGAGCGCGCTGCGCGGTGCGCGCGCGCGCGCGCGCG 3999
QY 1720 GCGCGCGTACCGCGCGCGCGGATGCTGCGTGTGCGCGCTGCTGCGCGCGCGCGCG 1779
Db 4000 GTGACCTCGGTGCGGCGCGCTTGGGAGGCGCGCGCTGCGCGCGCGCGCGCGCGCG 4059
QY 1780 GTCGAGGTCTTGTCAACCGCGCGCGCGCGCGCTGCTTCCGAGAGGTCCACTTCCCGCG 1839
Db 4060 GAGCGGTGCGCGCGCGCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 4119
QY 1840 GCGGACCGGGAATCTGCTCTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1892
Db 4120 GACACCGCGCTCGGCG 4172

RESULT 10

US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

```
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861.289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Query Match          3.9%; Score 78.2; DB 10; Length 11220;
Best Local Similarity 41.9%; Pred. No. 3.4e-09;
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;

Qy 469 GCCTCGCTCCGCTCTACACATGACGCCCCCGGCTGCTCTGAGCCACCCCAA 528
Db 2740 GACTCCCCCGCGTCCAGGGCTCCGTGAGGACTCTCTGGCGTACCGATCGACTGGA 2799
Qy 529 GCGCCGGTACACACCCAGCGGCTTACCAGCTGTACTACCTGCACTCCGACAGAAC 588
Db 2800 GCGCTCGGGTCCGAGCGCTCCAGCGCGCGCGGCTGTCCGGCGCTGGCTCGTGTG 2859
Qy 589 GCGCCCGCGGCTGGGACACCGGACGACGACCGGCGGTGCTTACGCGACACCGCG 648
Db 2860 GTCGCCGAGGACCGTTCGCCCGAGGCGCGCGGCTGTCCGGCGCTGTCCGGCGCG 2919
Qy 649 ACCGTGATCGCGCTCGGCGCGGCTTCCCGGTGCTGCGGTCGCGGCTCGCGGAC 708
Db 2920 GCGGACCCGCTACAGCTGGAGCTGTCCCGGTGCGGCGACCGGCTCGCGCGGACG 2979
Qy 709 GCGAACACGCG---AGGGTTCGGCGCGCGGCTGCTGCGGCTCGCGACCGCGGAC 765
Db 2980 CTGGCGAGGCGCTCGCGCGCGCGGCTGGAGCGCTGAGCGGCTCTCTGCTGCTCGG 3039
Qy 766 GACGCGCTCGGAGTACACGAGGACGATCTCTACTGCTGAGACGCGGCGGTTACG 825
Db 3040 TGGGACGAGAGCGCACCGCGCGGCGCGCGGCTTACCGCGGCGCGCGCGCGGAC 3099
Qy 826 TTCACGCGCGCTCGCGACCGCGCTATCTGTCACACCGGAGGCTCGCGCGCGCGG 885
Db 3100 CTCACCTGGTGGAGCGCTGGAGACCGCGCGGCTCGCGCGCGCGCTGTGTGCTGAC 3159
Qy 886 GCGGAGATCGAGACCGCGGAGTGTTCGCGACCGCGGAGTCCACTGGGACACCGCG 945
Db 3160 CACGCGCGGCTGCTGCTCGCGCGGCGCGGACGACGCTACCTCCCGCGCGCGCGG 3219
Qy 946 GGAGATGGGTCTGCGTCTCATCGGAGACTGCGGTTACGCGCGGTTCTACACCTCGCG 1005
Db 3220 TGGGCGATGGCGCGGCTCGCGCGCTTGAGCACCGCGGCTGGGCGGCGCTGATCGAC 3279
Qy 1006 CTGCGCGACTGGACACTTCGCGGCACTTCGACTACCGACCGACCGCGCTCGCGG 1065
Db 3280 CTGCGCGCTCGGACCGCGCGGCGCGGCTTGACCGCATGACACCGCTCTCGCGCG 3339
Qy 1066 GAGTCCCGCGGCTCTGAGATCACCGCGACGACGCGGACGCGCTGCGTGTGCTGCG 1125
Db 3340 ACGGTGAGGACGCTCGGCTAGCGCGCTCCGGGCTGCTCGCGCGCGCGCTCGCG 3399
Qy 1126 GCGAGCATGGAGCGCTTCCAGCGCGCGGCTCCCGATGAGTACGCTACTTGGACAGG 1185
Db 3400 GCGTCCCTCCGCGCGCGGCGCGGCTTTCGCGGTGGTGGCAGCGCGGCGGCGGCTG 3459
Qy 1186 TGGGACGCGGAGCACTTCCAGCGCGCGGCTCCCGCATGAGTACGCTACTTGGACAGG 1245
Db 3460 GTCACCGTGGCGAGGCGCTGCGCGCGCGGCGGCGGCGGCGGCTGCGCGCGGCG 3519
Qy 1246 GACTGTGACGCGGCGCTCACCTGGCGCATGATCGACGCGCGGAGACCAAGCGCTCG 1305
Db 3520 GCGGACACCTCTCTCTCCACACACCGCTTCCGCGCGGAGGCGCGCGGACGCGCTCC 3579
```

RESULT 11

```
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; CURRENT APPLICATION NUMBER: US/09/860.846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-5
```

```
Query Match          3.9%; Score 78.2; DB 9; Length 36778;
Best Local Similarity 41.9%; Pred. No. 3e-09;
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;

Qy 469 GCCTCGCTCCGCTCTACACATGACGCCCCCGGCTGCTCTGCGACCGCCCAA 528
Db 18427 GACTCCCCCGCGTCCAGGGCTCCGTGCGAGGACTCTTGGCGTACCGCATCGACTGGA 18486
Qy 529 GCGCGCGTACACACCGGCGGCTTACGAGCTGTACTACCTGCACTCGCGACGAGAAC 588
```

Db 18487 GCCTCGCGTCCCGACGCGTCCGAGCGCGCGGCTGTCCGGGCGCTGGCTCGTCGTC 18546
QY 589 GGCCTCGCGGCGTGGACACACGAGACAGACGAGCGCGTTCACGACACACGCGC 648
Db 18547 GTCCCGAGGACGCTTCGCGGAGCGCGCGCGTGTCTCGCGCGTGTCCGGCGCGG 18606
QY 649 ACCGTGATCGCGTGGCGCGCGCTTCCTCCCGTGTGGTTCGGGTCGCGTTCGCGAC 708
Db 18607 GCGGACCCGATACAGTGGACGTGTCCCGCTGGGAGACCGGAGCGGTTCGCGGACG 18666
QY 709 GCGAACACGCG---AGGTTGCGCGCGCGCGGCGTGTGGCGTTCGCGACACGCGAC 765
Db 18667 CTGGGCGAGCGCGTGTGGCGCGCGGTGGAGCGGTGAGCGGTGTCTCGTGTCTCGG 18726
QY 766 GACGCGTCCGCAAGTACAGGAGCAAGTACCTTACTGTGTGACCGACGCGGCTTCAC 825
Db 18727 TGGGACGAGCGCGACCCCGCGCACCCCGCGCTTCACCGGGGACCGGCGCGAC 18786
QY 826 TTCACCGCGTCCCGACCGCGTATCGTAAACACGAGCGGTGGCGCGCGCGCGCGCC 885
Db 18787 CTACCGTCCCGACCGCGTCCGAGCGCGCGCGTGTGGCGCGCGCGCGCGCGTAC 18846
QY 589 GGCCTCGCGGCGTGGACACACGAGACAGACGAGCGCGTTCACGACACACGCGC 648
Db 18547 GTCCCGAGGACGCTTCGCGGAGCGCGCGCGTGTCTCGCGCGTGTCCGGCGCGG 18606
QY 649 ACCGTGATCGCGTGGCGCGCGCTTCCTCCCGTGTGGTTCGGGTCGCGTTCGCGAC 708
Db 18607 GCGGACCCGATACAGTGGACGTGTCCCGCTGGGAGACCGGAGCGGTTCGCGGACG 18666
QY 709 GCGAACACGCG---AGGTTGCGCGCGCGCGGCGTGTGGCGTTCGCGACACGCGAC 765
Db 18667 CTGGGCGAGCGCGTGTGGCGCGCGGTGGAGCGGTGAGCGGTGTCTCGTGTCTCGG 18726
QY 766 GACGCGTCCGCAAGTACAGGAGCAAGTACCTTACTGTGTGACCGACGCGGCTTCAC 825
Db 18727 TGGGACGAGCGCGACCCCGCGCACCCCGCGCTTCACCGGGGACCGGCGCGAC 18786
QY 826 TTCACCGCGTCCCGACCGCGTATCGTAAACACGAGCGGTGGCGCGCGCGCGCGCC 885
Db 18787 CTACCGTCCCGACCGCGTCCGAGCGCGCGCGTGTGGCGCGCGCGCGCGTAC 18846
QY 1126 GCAGATGAGCGCTTACGCGCTCGCGCTTCGCGTACGCTAGCGTTCGCGTTCGCGGAC 1185
Db 19087 GCCTCCCTCCCGCGGACGCGGACGCTTCGCGGTGGTGGACGCGGCGCGCTGTCG 19146
QY 1186 TGGGAGCGGAGAGTTCCACGCGGACGCTTCACCGCGAATGGTTCGAGTGGGCTGG 1245
Db 19147 GTACCGGTGCGGAGAGCTTGGCGCGCGGAGCGGCGGCGGCTGGCGCGGAGCGC 19206
QY 1246 GACTGTAGCGCGGTACCTTGGCGCATCGAGCGCGCGGCGGAGCGGCGGCTGGCG 1305
Db 19207 GCGGACACCTCTCTCTCACACACCGCGCTTCGCGAGCGGAGCGGCGGAGCGACCT 19266
QY 1306 ATCGGTGATGAACACTGGAAGTACGCGCGGACGCGGCTTCGCGACGCGATCCGAC 1365
Db 19267 GTGCGCGCGGAGACTTCGCGCTTCGCGGCTTCGCGGAGCTTCGCGGAGCTGGCGG 19326
QY 1366 GGTACAGCGGAGAGTTCGATTCGCGGAGCTTCGCGGCTTCGCGGAGCTTCGCGG 1419
Db 19327 ACGGGACCGGTGTGACCTTCGCGGAGCTTCGCGGAGCGGCGGCGGCTTCGCG 19386
QY 1420 GCGGCTGTACACCTCTGTGAGACCGCGGCGGAGCGGCTTCGCGGAGCTTCGCGG 1479
Db 19387 GCGGCTTCGAGCGGACCGGCTTCGCGGAGCTTCGCGGAGCTTCGCGGAGCTTCG 19446
QY 1480 ACCACACACTTCGCGGAGCGGCTTCGAGCGGAGCGGCTTCGCGGAGCGGAGCG 1539
Db 19447 GAGCGCTCGCGGAGCGGAGCGGCTTCGCGGAGCTTCGCGGAGCGGAGCGGAGCG 19506
QY 1540 GCATACGAGATGAGTGTGAGATCGCGTGGGAGCGGCGGAGCGGAGCGGAGCGG 1599
Db 19507 GCGGCGCTCCACTGAGCGGCTTCGCGGAGCGGCGGCTTCGCGGAGCGGAGCGG 19566
QY 1600 GCGCGCTCCCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1659

Db 19567 GTCTCGGTCTCTTCT 19626
QY 1660 GTCTCGAGGAGCGCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1719
Db 19627 GCGCGCGGTACGCGCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 19686
QY 1720 GCGCGCGATCGAGCG 1779
Db 19687 GTGACCTCGGTGCGTGGAGCGCGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 19746
QY 1780 GTCTGAGGTCTTCT 1839
Db 19747 GAGCGGTCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19806
QY 1840 GCGGACACCGGATCT 1892
Db 19807 GACACCGCGCTCGCTG 19859
RESULT 12
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match 3.9%; Score 78.2; DB 10; Length 36778;
Best Local Similarity 41.9%; Pred. No. 3e-09;
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;

QY 469 GCTCGCTCCGTCGCTTACCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 528
Db 18427 GACTCCCCCGCTCCAGGCTCCGTGAGGACTCTCTGCGCTACCGCATCGGTAAG 18486
QY 529 GCGCGCGGTCCACCG 588
Db 18487 CGCTTCGCGGTCCG 18546
QY 589 GCGCGCGCGGTGGAGACACCGGAGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCG 648
Db 18547 GTCCCGGAGGACGTTCCCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18606
QY 649 ACCGTGATCGCGTGGCGCGCGCGCTTCCTCCCGTGTGGTTCGGGTCGCGGCTCT 708
Db 18607 GCGGACCGCGTACGATGAGCGGTTCCTCCCGTGGCGAGCGCGCGCGCGCGCG 18666
QY 709 GCGAACACGCG---AGGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 765
Db 18667 CTGGGCGAGCGCGTGGCGCGCGCGCGTGGAGCGCGTTCGCGGCGCTCTCTCTCT 18726
QY 766 GAGCGGTTCGCAAGTACAGGAGCGAGTACCTTACTGTGTGAGCGGCGGCGGTTAC 825
Db 18727 TGGGAGGAGGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18786
QY 826 TTCACCGCGTCCCGGAGCGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 885
Db 18787 CTACCGTCCCGGAGCGCGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18846

Qy	1186	TGGGACGCGAGCAGTTCCACGCGCAGACACCTCACCCGCGAATGGCTCGACTGGGGCTGG	1245
Db	20317	GTACACCGGTCCGAGAGACCTTCGGCGCCGCGAGCGCAGCCGCGTGGCCGCGAGCGC	20376
Qy	1246	GACTGGTACGCGGCGCTCACCTGGCCATCGACTGACGCGCCCGAGACAACGCGCTCGCC	1305
Db	20377	GCCGACACCTCTCTCTCCACACCAACCCCTCCGGCAGCGAAGCGCGGACACCTCC	20436
Qy	1306	ATCGCGTGGATGAAACAATGGAAGTAGCGCGACGCGAGCTCCCAACCGACGATCCGAC	1365
Db	20437	GGTGGCGCGAGGAGACTCCGGCTCGCCGGGCTCGTGC CGGAACCTCGGAGACTGGGCGG	20496
Qy	1366	GGCTACAAACGGCAGAACTCGTCTCGCGAGCTCGGCTCGCCCGACAGC-----CT	1419
Db	20497	ACGGCCACCGCTGTGACTCGGACCTCAGGACGGAGGCGCGCCGCGCTGCTCGCC	20556
Qy	1420	GGCGGCTGGTACACCCCTCTTGAGCACCCCGCTGGCAGCTGAGCGAACTACGTCAACGCC	1479
Db	20557	GGCGTCTCCGACGCGCACCCGCTCAGCGCGCTCTCCACCTGCGCGCCACCGTCTGACTCC	20616
Qy	1480	ACCACCACATCCCGACCGGACCGTGCAGGCGAGCGCGCTCTGCGCATGGAACGGAGCG	1539
Db	20617	GAGCGCTCGCGCGGACCGAGCGGAGCGCGTCCGCGCTGTCTGAGCGGAAGCGCCACC	20676
Qy	1540	GCATACGAGATCGAGCTCGACATCGCTGGGACACCGCAGCAGACGTCGGCATCTCGGTG	1599
Db	20677	GCCGCGCTCCACTGGACCGGCTCTCTGGGAGCGCGGGTGC CGGAGGCCGTCCGCCCC	20736
Qy	1600	GGCGCGTCCCGGACGGAACCCGCGACACGAACTCGGCAAGTACGAGCAGACCTGTAC	1659
Db	20737	GTCTTGGTCTCTTCTCTCTCGGTCCGCGCATCTGGGGCGCGCGGTTCAGGGCGCGTAC	20796
Qy	1660	GTCCACGAGGACCCCTCGGACCTCGCGGGTACTCGCTCGCCCGCTACTCCGAGCGCGC	1719
Db	20797	GCCCGCGTACGGCTTCTCTGACGCGCTCTGCCGGTCAGACCGGGCGGACGGCCCCCACC	20856
Qy	1720	GCCCCCATCGACCCCGCGCCGATCGGTGCACCTGGCGCATCTCTCTCGACACCCAGAGC	1779
Db	20857	GTGACCTCGGTGGCTTGAGGCCCTTGGAGGGCAGCGCGCTACCGAGGGTGCACCGGG	20916
Qy	1780	GTCGAGGTCTTGTCAACGGCGGCGACACCGTGTCTCCACGAGGTCCTACTTCGCCGAG	1839
Db	20917	GAGCGGTGCGCGGCTCTGGCTCGGCGCCCTCTGCCCGCGGACGGCGCTACCGCCCTG	20976
Qy	1840	GGCGACACGGGAATCTCTGCTCTATACGACGAGCGGCGGCCCGCACACTTCACCGG	1892
Db	20977	GACACCGGCTCGCGCACGGGACACCGCGTCTACGATCGCGACGCTCGACTG	21029

Query Match	3.9%	Score 77.6	DB 9	Length 12441
Best Local Similarity	42.9%	Pred. No. 4.6e-09		
Matches 680	Conservative 0	Mismatches 879	Indels 25	Gaps 5
QY 367	CGGGCCATCTCACGCCGCGCGTGTCTCAGGAGCGCGCGGAGCAGCTCGCCCTCATC	426		
DB 11836	CGGACACCGGACACTCGACCCGACAGAGTGGCGCGCGGTTCACACCCGACCTCG	11777		
QY 427	TTGCGCGTGTGTGCGCCCTGACGCGCGGCATCGCTCCGGCTCGCTCCGTGCCGTG	486		
DB 11776	GCGTGTGCGGCTCCACTCTGGGCGCGCCCTGCGCGCGGACGACTCGGGAAGTCT	11717		
QY 487	TACCACATGACGCCCCCAGCGGTGTCTGACACCCCAACCGCGGTACACACCCAC	546		
DB 11716	GCGGACGAGCAGGCGCTCGGCTGTACTTTCAGCGCGGACGCGCTCGGCTGCGGTG	11657		
QY 547	GCGGCTTACAGCTGTACTACTCTGCATCTCCACAGACACAGCGCCCGGCGGTGGAC	606		
DB 11656	GACGGCGCGCGCGCGGAGCTCTGGGACGCGGAGGTCTTACGTTTCCAGCGCCACAAG	11597		
QY 607	CACGGACACGACGCGGCGTCCCTTTCACGACACGACGCGCTGATGCCGCTCGG	666		
DB 11596	GCGGTACAGGCTTTCGAGGCGGCGCGTGTCTACGACGACGCGGACTGCGCGCGCG	11537		
QY 667	CCGACTTTCGCGGTGTGTCGCGGCTCGGCGACCGGACACGCGAGGTTTC	726		
DB 11536	ATCGCGGCTTCCACAACTTCGGCTTCGACTGCGCGGCGGACGCGCGGCGGAC	11477		
QY 727	GCGCGCGCGCGGTGTGTCGCGCTCGGACCCAGCGGACGACGCGGTTCGGAAGTAC	786		
DB 11476	AACGCCAAGATGACGAGGCGCGCGCGCATG-----GGCCTCACTCCCTC	11429		
QY 787	GAGCAGTACTCTACTGTGACGACGCGGCGTTCAGTTCAACGCGCTCGCCGACCC	846		
DB 11428	GACGGTTTCCGAGGTATCGACGGAACCGGCGCAACACGCGGCTTACGCGGAC	11369		
QY 847	GTCTCTGTCAAACCGACGCGTTCGCGCGCCACCAACGCGCGGATCGAGAAGCGG	906		
DB 11368	CTCGGAGCTCCCGCGCTCTCTGTCGCGACACGACGCGGCTCAACAACAC	11309		
QY 907	TGGTTCGCGACCCAGATCCATCGGACACCGCGCGGAGATGGGTCTCGGCTATC	966		
DB 11308	CAGTACGTGATCGTCTGAGATGACGAGGCGCACCGGCGATCCACCGGACTCTG	11249		
QY 967	GGACGACTTCCGGTACGCGCGTTCACACTGCGCGAACCT-----GCGCGACTGG	1020		
DB 11248	GAGTCTCTGAGCGGAGGCGTGCACACCGCGGCTACTCTCGCGGCTCGCAGG	11189		
QY 1021	CTTCGCGCAATTCGACTACCCGAAACAGCGCTCGCGGCGATCTGAGTCCGCCGAC	1080		
DB 11188	CTGAGCGGTACCGCGGCGACGCGACGCGCGGTGCGCGACACGGAACGCTCGCGG	11129		
QY 1081	TTGAGATCACCAGAGGAGGACAGGCACTGGGTGTGCGCGGCGGACGATGGAGCC	11140		
DB 11128	CGCGTGTGTCTCTGCGGACCGGCGACCGCATCGGCGACACGACATCCCGGCGT	11069		
QY 1141	TACGGCATCGGCTCCCATGACGTACGCTACTTGGACAGGCACTTGGGAGCGGAGC	1200		
DB 11068	GACCTGTGCTGTCTGCGGCGCGCGCGGAGACTGACCGGCGCCACCGGAGACG	11009		
QY 1201	TTCCACGCGGACGCTACCCCGCAATGGCTCGACTTGGGCTGGGACTGGTACGCGG	1260		
DB 11008	GCCCCGCGC-CCGCTCGGCGCCCCAGACATCCAGCCCACTTGGACGCTCCGATG	10950		
QY 1261	GTCACCTTGGCCATCGATCGAGGCGCGCGAGACAGCGGCTCGCATCGGCTGGATG	1320		
DB 10949	ACCGCGCGCGCTTTCGCGCACCGCGCGCGGAGGCTGCGCGCACCCCGGAGCGG	10890		
QY 1321	AACGTGAAGTACGCGGACGCGAGTCCCGACGACGATCCCGACGCTACACGGGAG	1380		
DB 10889	CTGGGCGGCGGTTCACGCGGTGCGGCGAGACCTCTCGCGCGCGGCGCTCTGTGCG	10830		
QY 1381	AACCTCATGCTGCG- --CGAGCTGCGGCTCGCCCGACAGCCTCGGCGGCTGGTAC	1437		


```
Db 10829 GACGAGGCGGGAACACCGCCGACCTCGTCCGGCTCGCGGTACGGCAACAGC 10770
QY 1438 CTGAGCACCCTCGGTGAGCGCTGACGAACTACGTACCGCCACACACACATCTCCCGAC 1497
Db 10769 CCGTTACCCCGCTGGAGAGCGCCGACGACCTGGGGGTGACGGGACCCCTTCGG 10710
QY 1498 CGGACCGTGGAGCGCGGCTGCTGTCATGGAACGACGCGCATACGATCGAGCTC 1557
Db 10709 CGCCTCCCTCGCCCTTTCGGGAGGTCCCGAGTCCGACCGCGGTCCGAGACCGCC 10650
QY 1558 GACATCGCTGGGACACCGCGAGAACGTTCGGCATCTCGGTGGCGCTCCCGCGAGGA 1617
Db 10649 GCGGGGGGTACTGGAAGAACACCTGCTCCGGCTCGAAGCAGCGGGGTCTTCGACGG 10590
QY 1618 ACCCGGCAACAGAA---CATCGGCAAGTACGGAGCAGACCTGTACGTTCGACCGAGACCC 1674
Db 10589 GCGCTCGCGAGAACCGCTTTCGGGTACAGCTGCGGCTTACCCGCGCGGACCTCGC 10530
QY 1675 TCGACCTCGCGGGGTACTCGCTCGCCCGCTACTCGAGCGCGCGCGCCCATCGACCC 1734
Db 10529 ATGTTCCGCTGCACCTTCTGCGTCCGTGTGACCGGCGCGCTACGACCGGTCCGCGCTC 10470
QY 1735 GCGCGCCGATCGTCACCTCGCATCTCGTCGACACCCAGAGCTCGAGTCTTCGTC 1794
Db 10469 GACGCGGCAAGCGCATGTTCGGTTCGTCATCGACGAGATACCGCGGGCAACCCCTCG 10410
QY 1795 AACGCGGCGCACACCGTCTCCCGACGAGGTCCACTTCGCGGAGGGGACACGGAATC 1854
Db 10409 GCGATGTACTTTCGCGGGCTTGAGCGCTACCAACCCGCGCTCGGAGCTGGCC 10350
QY 1855 TCGCTTACACGAGCGGGCGCCCGACACTTCACCGCATCGTCGCGGAGATGGC 1914
Db 10349 GCGCACGCGACGACCGGCTCGCGCCACCGCTACACGAATCTCTTCGCGCTCAC 10290
QY 1915 CAGGCGATCTAGCGATGCACACC 1938
Db 10289 GAGCGACCCCTGGAGCGCAGCCC 10266
```

RESULT 15

US-09-860-846-3/C

; Sequence 3, Application US/09860846

; Patent No. US20020164742A1

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.438US1

; CURRENT APPLICATION NUMBER: US/09/860,846

; PRIOR FILING DATE: 2001-05-18

; PRIOR APPLICATION NUMBER: 09/105,537

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 3

; LENGTH: 13613

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-860-846-3

```
Query Match          3.9%; Score 77.6; DB 9; Length 13613;
Best Local Similarity 42.9%; Pred. No. 4.5e-09;
Matches 680; Conservative 0; Mismatches 879; Indels 25; Gaps 5;
```

```
QY 367 CGGGCATCTCACGCGCGCGGTGCTCCAGGAGCGCGCGGAGCACTGCGCCCTGATC 426
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12523 CGGACACCGCAACTCGACCCGACGAGGTGGCGCGCGGTTCACACCCCGACCTCG 12464
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 427 TTCGGCGGTGTGTGCCGCTGCAAGCCCGGCGATCCGCTCGGGGTCTCGTCCGCTC 486
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 12463 GCCGTGTCGGGTCCAACTCTTGGGCGCGCCCTCGCGCGCGACGAGTCGCGAAGTTC 12404
QY 487 TACCACATGACGCCCCAGCGGCTGGGTCTCGGACCCCAACGCGCGGTACCAACCCAC 546
Db 12403 GCCGAGGACGCGGCTTGGGCTGTACTTCGACGCGCGGACGCGCTCGGTTCGGGGTC 12344
QY 547 GCGGCTTACCAGCTGTACTTACCTCCGACACCAAGAACGCGCGCGCGGTGGGAC 606
Db 12343 GACGCGCGCGCGCGGAGCCTCGCGGACGCGGAGGTTCAGCTTCACGCCACCAAG 12284
QY 607 CACGCGAGCACGACGCGGCTCGCCTTCACGACACCGGACCGGTGATCGCGTGGG 666
Db 12283 GCCGTCAACGCGCTTTCGAGGGCGCGGCTGTACGACGCGGAGCTCCCGCGCGG 12224
QY 667 CCCGACTTCCCGGTGTGTCGCGGTCGCGGTCGTCGCGCACCGCGAACACGACGAGGTTC 726
Db 12223 ATCCGCGGCTTCCAACTTTCGCTTCGACCTGCGCGGCGGACGCGCGCGGAC 12164
QY 727 GCGCGCGGCGGCTCGTTCGCGCTCGGACACCGACCGGAGCGGTTCGGAAGTACAC 786
Db 12163 AACGCAAGATGAGGCGGCGCGCGCGCATG-----GGCCTCACCTCCCTC 12116
QY 787 GAGCAGTACCTCTACTGTGACGACGCGGCTTACGTTTACCGCGCTCCCGACCC 846
Db 12115 GACGCTTTCGAGGTTCATGACCGGAACCGCGGCAACCGCGCTTACCGCGAGCAC 12056
QY 847 GTCATCGTCAACACGAGCGGTGCGCGCGCACCGCGCGGAGATCGAGAACCGCGAG 906
Db 12055 CTCGCGGACCTCCCGCGGCTCTCTGTCGCGACACGACCGCGCGCTCAACAACAC 11996
QY 907 TGGTTTCGCGGACCCCAAGATTCATCTGGGACACCGCGCGGAGATGGTCTGCGTATC 966
Db 11995 CAGTACGTGATCGTGGAGATCGACGAGGCGCACCGGCGATCCACCGCGACCTCTCATG 11936
QY 967 GGACGACTCGGTAGCGCGGTCTTACACTCGCGCAACT-----GGCGACTGGACA 1020
Db 11935 GAGTCTCTAAGCGGCAAGCGGTGCACACCGCGCTTCTTCGCGGCGGTGCGACAG 11876
QY 1021 CTTTCGCGCACTTCGACTTACCCGAACACGCGCTCGGGCGCATGAGTGCOCGACCTG 1080
Db 11875 CTGGAGCGGTACCGCGGCGCGCGCGCGCGCGCTGCGCGCACACCGCGCTCGCGCG 11816
QY 1081 TTCGAGATCACCGGACGACGCGGACACGCGCTCTGGTGTCTGCGCGCACGATGACGCG 1140
Db 11815 CGCGTGTCTCTCTGCGCGCGCGCGCGCGCATGCGCGGACGACATCGCGCGGTTCGCG 11756
QY 1141 TACGCGATCGGCTCCCGATGACGTACGCTTCTGGACGACCGCTGGGACGCGCGAGCAG 1200
Db 11755 GACCTGTGCTCTCTGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 11696
QY 1201 TTCCACGCGGAGCGCTCACCCCGCAATGGTCTCGACTGGGGCTGGGACTGTGTACCCCTC 1260
Db 11695 GCGCGCGCGC-CGCTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATG 11637
QY 1261 GTCACCTGCGCATGATGACGCGCGCGGACACGCGCTCGCGATCGCGTGGATGAC 1320
Db 11636 ACCGCGCGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAT 11577
QY 1321 RACTGGAAGTACGCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1380
Db 11576 CTGGGGGCGCGGTCCACGCGCTGCGCGGACACCGCTGCGCGCGCGCGCGCGCGCGCG 11517
QY 1381 RACTGATCTGTCG---CGAGCTGGGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 1437
Db 11516 GACGAGCGCGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11457
QY 1438 CTGAGCACCGCGGTGCGCGCGCTGACGAATACGTACCGCGCGCGCGCGCGCGCGCGCG 1497
Db 11456 CCGTTACCGCGGTGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11397
QY 1498 CGGACCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1557
Db 11396 CGCCTCTCGCGCTGTTCGGGCGAGTTCGCGGAGTCCGCGCGCGCGCGCGCGCGCGCG 11337
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:23:32 ; Search time 34.9327 seconds
(without alignments)
1979.723 Million cell updates/sec

Title: US-09-868-328B-4

Perfect score: 2837

Sequence: 1 MTPAIRRAVLQAGAGALA.....LYTGGPAHFTGIVVREIGQ 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2837	100.0	521	22	AA82301
2	2352	82.9	517	20	AA82301
3	2352	82.9	517	23	AA82301
4	2265	79.8	484	20	AA82301
5	2265	79.8	485	23	AA82301
6	499.5	17.6	943	21	AA82301
7	468.5	16.5	1277	21	AA82301
8	461	16.2	923	21	AA82301
9	427	15.1	516	21	AA82301
10	416.5	14.7	533	19	AA82301

11	389.5	13.7	490	16	AA82301
12	386	13.6	750	20	AA82301
13	386	13.6	750	20	AA82301
14	366	12.9	556	15	AA82301
15	359.5	12.7	532	5	AA82301
16	359.5	12.7	532	12	AA82301
17	348	12.3	581	20	AA82301
18	348	12.3	581	20	AA82301
19	348	12.3	581	21	AA82301
20	292.5	10.3	671	23	AA82301
21	277.5	9.8	636	16	AA82301
22	277.5	9.8	636	21	AA82301
23	274.5	9.7	636	13	AA82301
24	274.5	9.7	636	14	AA82301
25	266	9.4	583	22	AA82301
26	266	9.4	648	23	AA82301
27	261.5	9.2	1116	20	AA82301
28	258.5	9.1	645	23	AA82301
29	244.5	8.6	479	23	AA82301
30	243	8.6	670	21	AA82301
31	241	8.5	553	15	AA82301
32	241	8.5	553	16	AA82301
33	239.5	8.4	1487	16	AA82301
34	235.5	8.3	492	23	AA82301
35	235.5	8.3	591	21	AA82301
36	231.5	8.2	587	23	AA82301
37	223.5	7.9	569	23	AA82301
38	222.5	7.8	554	23	AA82301
39	222	7.8	591	23	AA82301
40	220	7.8	581	23	AA82301
41	220	7.8	588	21	AA82301
42	205.5	7.2	630	17	AA82301
43	205.5	7.2	637	19	AA82301
44	202.5	7.1	429	22	AA82301
45	202.5	7.1	433	22	AA82301

ALIGNMENTS

RESULT 1
AA82301
ID AA82301 standard; Protein; 521 AA.
AC AA82301;
DT 09-JUL-2001 (first entry)
DE Arthrobacter ureafaciens levan fructotransferase.
KW Levan fructotransferase; difructose dianhydride IV; sweetener.
OS Arthrobacter ureafaciens.
FH Key Location/Qualifiers
FT Peptide 1..33
FT Protein 34..521
FT /label= Signal_peptide
FT /label= Mature_protein
XX WO200129185-A1.
XX 26-APR-2001.
XX 19-OCT-2000; 2000WO-KR01183.
XX 19-OCT-1999; 99KR-0045302.
XX (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
XX (REAL-) REALBIOTECH LTD.
XX Rhee S, Song K, Kim C, Ryu E, Lee Y;

Penicillium purpur
Bacillus sp. L7 en
Bacillus L7 endo-1
Inulinase preprote
Saccharomycs cere
Preinvertase. Sac
Schizosaccharomyce
S. pombe invertase
Schizosaccharomyce
Fructosyl transfer
Tomato plant inver
Wild-type tomato i
Tomato vacuolar in
Tomato acid invert
Lycopersicon penne
Fructosyl transfer
S. rochei strain E
Amino acid sequenc
Streptococcus poly
Corn invertase pro
Tomato acid invert
Tomato acid invert
Bacillus circulans
Staphylococcus epi
Arabidopsis thalia
Arabidopsis thalia
Herbicidally activ
Herbicidally activ
Fructosyl transfer
Herbicidally activ
Herbicidally activ
Arabidopsis thalia
1-Sucrose:sucrose
Artichoke sucrose
Corynebacterium th
C glutamicum prote

DR WPI; 2001-308483/32.
 XX N-PSDB; AAF30918, AAF30919.
 XX Producing difructose dianhydride IV from sucrose, involves reacting
 PT sugar solution in the presence of levansucrase to produce levan, and
 PT reacting levan solution in the presence of levan fructotransferase to
 PT produce DFA IV.
 XX Claim 2; Page 47; 72pp; English.
 XX The present sequence is that of Arthrobacter ureafaciens K2032
 CC levan fructotransferase (see AAF30918). The enzyme can be
 CC obtained by cultivation of claimed Escherichia coli JUD81
 CC (KCTC 0877BP), which carries claimed expression vector pUDFA81
 CC comprising the levan fructotransferase gene (see AAF30918). A
 CC claimed process for producing difructose dianhydride IV from
 CC sucrose comprises subjecting sugar solution to reaction at room
 CC temperature or lower in acidic buffer of pH 3.0-7.0 in the presence
 CC of a levansucrase derived from Zymomonas mobilis to produce levan,
 CC purifying the levan from the reaction solution, and subjecting it
 CC to reaction at 25-30 degree C for 3-10 hours in acidic buffer of pH
 CC 3.0-7.0 in the presence of levan fructotransferase, preferably
 CC obtained from E. coli JUD81. The product is useful as a low-calorie
 CC sweetener.
 XX Sequence 521 AA;
 SQ Query Match 100.0%; Score 2837; DB 22; Length 521;
 Best Local Similarity 100.0%; Pred. No. 1.2e-240;
 Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTPAISRRVAVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDPQRPVT 60
 DB 1 MTPAISRRVAVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDPQRPVT 60
 QY 61 THGAYOLYLHSDQNGPGGWDHASTTGDGVAFTTHGTVMPLRDPDFPVWGSAAVVGTA 120
 DB 61 THGAYOLYLHSDQNGPGGWDHASTTGDGVAFTTHGTVMPLRDPDFPVWGSAAVVGTA 120
 QY 121 GFCAGAVVALATOPTDGVVRKYQOYLWSTDDGFTFTALPDPIVNTDGRAATTAEIEN 180
 DB 121 GFCAGAVVALATOPTDGVVRKYQOYLWSTDDGFTFTALPDPIVNTDGRAATTAEIEN 180
 QY 181 AEWFRDPKIHWDTARGEWCVIGRLRYAAYFTSPNLRDWTLRNFDYPNHALGGIECPDL 240
 DB 181 AEWFRDPKIHWDTARGEWCVIGRLRYAAYFTSPNLRDWTLRNFDYPNHALGGIECPDL 240
 QY 241 FEITADDGTRHWVLAASMDAYGIGLPMYAYWTGTWGDGEQFHADDLTPOWLDGWDWYAA 300
 DB 241 FEITADDGTRHWVLAASMDAYGIGLPMYAYWTGTWGDGEQFHADDLTPOWLDGWDWYAA 300
 QY 301 VTWPSIDAPETKRLATAMNNNKYAAARDVPTDASDCYNGQNSIVRELRLARQPGGWYTL 360
 DB 301 VTWPSIDAPETKRLATAMNNNKYAAARDVPTDASDCYNGQNSIVRELRLARQPGGWYTL 360
 QY 361 STPVAALTNYVTATTPLPDRTVGSAVLPWNGRAYEIEIDIAWDTATNNGISVGRSPDGT 420
 DB 361 STPVAALTNYVTATTPLPDRTVGSAVLPWNGRAYEIEIDIAWDTATNNGISVGRSPDGT 420
 QY 421 RHTNICKYGADLYVDRGSDLAGYSLAPYSRAAAPIDPGARSVHLRLVDTOSVEVFVNA 480
 DB 421 RHTNICKYGADLYVDRGSDLAGYSLAPYSRAAAPIDPGARSVHLRLVDTOSVEVFVNA 480
 QY 481 GHTVLSQQVHFAEGDGTGSLYTDGGPAHFTGIVVREIG 519
 DB 481 GHTVLSQQVHFAEGDGTGSLYTDGGPAHFTGIVVREIG 519
 RESULT 2
 AAY04105
 ID AAY04105 standard; Protein; 517 AA.
 XX
 AC AAY04105;

XX 10-JUN-1999 (first entry)
 XX Arthrobacter nicotinovorans levan fructotransferase prote:n #2.
 DE Arthrobacter nicotinovorans; levan fructotransferase.
 KW Arthrobacter nicotinovorans; levan fructotransferase.
 XX OS
 XX Arthrobacter nicotinovorans.
 PN JP11069978-A.
 XX 16-MAR-1999.
 PD 28-AUG-1997; 97JP-0232421.
 PF 28-AUG-1997; 97JP-0232421.
 PR 28-AUG-1997; 97JP-0232421.
 XX (NIOC) NIPPON OIL CO LTD.
 PA WPI; 1999-247463/21.
 XX N-PSDB; AAX19827.
 DR Levan fructotransferase gene - for recombinant production of levan
 XX fructotransferase
 PT Claim 3; Page 8-9; 14pp; Japanese.
 XX The present sequence represents Arthrobacter nicotinovorans levan
 CC fructotransferase. The present invention also describes a method
 CC for the preparation of levan fructotransferase in which a transformant
 CC is cultured in a medium and levan fructotransferase is collected from
 CC the culture. The method can prepare levan fructotransferase in a
 CC large amount.
 XX Sequence 517 AA;
 SQ Query Match 82.9%; Score 2352; DB 20; Length 517;
 Best Local Similarity 81.3%; Pred. No. 5.1e-198;
 Matches 421; Conservative 41; Mismatches 52; Indels 4; Gaps 2;
 QY 1 MTPAISRRVAVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDPQRPVT 60
 DB 1 MYDISRRALQAGAGALALFMSNAIPVAHAQA--SLRAIYHMTPPSGWLCDPQRPVH 58
 QY 61 THGAYOLYLHSDQNGPGGWDHASTTGDGVAFTTHGTVMPLRDPDFPVWGSAAVVGTA 120
 DB 59 THGAYOLYLHSDQNGPGGWDHASTTGDGVSYTHGVAWMPQDPFVWGSAAVVDTA 118
 QY 121 GFCAGAVVALATOPTDGVVRKYQOYLWSTDDGFTFTALPDPIVNTDGRAATTAEIEN 180
 DB 119 GFCAGAVVALATOPTDGVVRKYQOYLWSTDDGFTFTALPDPIVNTDGRAATTAEIEN 176
 QY 181 AEWFRDPKIHWDTARGEWCVIGRLRYAAYFTSPNLRDWTLRNFDYPNHALGGIECPDL 240
 DB 177 AEWFRDPKIHWDATRNEWVCVIGRLRYAAYFTSPNLRDQWKSNDYFNHALGGIECPDL 236
 QY 241 FEITADDGTRHWVLAASMDAYGIGLPMYAYWTGTWGDGEQFHADDLTPOWLDGWDWYAA 300
 DB 237 FEMTAGDGTGRHWVFGASMDAYSIGLPMYAYWTGTWGSNGTAFADNLTPQWLDGWDWYAA 296
 QY 301 VTWPSIDAPETKRLATAMNNNKYAAARDVPTDASDCYNGQNSIVRELRLARQPGGWYTL 360
 DB 297 VTWPAVEAPETKRLATAMNNNKYAAARNYPTDASDCYNGQNSITRELRLEROSGGWYTL 356
 QY 361 STPVAALTNYVTATTPLPDRTVGSAVLPWNGRAYEIEIDIAWDTATNNGISVGRSPDGT 420
 DB 357 STPVPALSNYATSSITPLPDRTVNGSVFLPWSGRAYELEIDISWDTAANVGVSGRSSDGS 416
 QY 421 RHTNICKYGADLYVDRGSDLAGYSLAPYSRAAAPIDPGARSVHLRLVDTOSVEVFVNA 480
 DB 417 RHTNICKYGADLYVDRASSEQSYALAPYTRAAAPIDANARSVHLRIFVDTSQVEVFVNS 476
 QY 481 GHTVLSQQVHFAEGDGTGSLYTDGGPAHFTGIVVREIG 518

CC and P. macerans, respectively, which have 2,6-beta-D-fructan hydrolase
 CC activity. These are preferred enzymes of the invention. The enzymes
 CC can be used to remove microbial biofilm from surfaces, particularly for
 CC reduction of biocorrosion, for treatment of dental plaque and for
 CC treatment of multiple sclerosis (eliminating biofilm from the lungs; to
 CC hydrolyze slime; in detergents and for production of fructose oligomers
 CC from levan or phleum (useful as sweeteners or for conversion to
 CC fructose). The nucleic acids that encode the enzymes are useful for
 CC recombinant production and as source of probes and primers for
 CC identification and cloning of related fructan hydrolases (all claimed).
 XX
 SQ Sequence 943 AA;

Query Match 17.6%; Score 499.5; DB 21; Length 943;
 Best Local Similarity 30.8%; Pred. No. 1e-34;
 Matches 157; Conservative 67; Mismatches 223; Indels 63; Gaps 19;
 QY 40 RAVYHMTPPSGWLCDDPQRYTHGAYQLYLHSDQNGPGWGHASHTTGDVAFTHHGTVM 99
 Db 383 RPDYHSPARGSADPNGLVYEGEYHLF--HOD---GGTWAHAVSTDLVHWRRLPIAL 436
 QY 100 PLRPDPFVWGSAGVGTANTAGF-----GAGAVVALATQPTDGVKRYQEQVLYWSTGGF 154
 Db 437 PWNDLGHVWGSADVLHNASGLFADSGGGLIAYYSYNPDENGQRIGLAYSKDRGR 496
 QY 155 TFT-ALPDPPIVNTDGRAATTPAEIENAEW-FRDPKIHMDTARGEWCVIGRLRYAAYT 212
 Db 497 TWYEAERPIVIEPNKGQDDP-----GGWDFRDPKVVRODEHNRVWVSGGDHIFRFT 551
 QY 213 SPNLRDWTLRNFDPYHNLGGI-ECDFLFEITADD-GTRHWVLAASMDAYGIGLPMTYA 270
 Db 552 STNLIDMTLTDSPGAYVGVGWECPDLFLQALVDDTGERKWLMISTGAN---PNTQG 607
 QY 271 YWCTWDGEQFHADDLTPQW-----LDWGDWVYAAVTWPSIDAPETKRLAIA 317
 Db 608 -----SAAEYFTEGELPEKGFVNDNPAGKVLATDYKEYIASMSFAGM--PDGRRVWLA 659
 QY 318 WMNNKYAARDVPTDASDYGNGONSIVRELRLARQPGGWYTLTSTPVAALTN-----YVTA 373
 Db 660 WMTNWDYPPFA-FPT---EGMKVLSIPRELTQLTKDQG-IRLAQTPIRELESRLGQLLFA 714
 QY 374 TTTLPDRTVG---SAVLPWNGRAYEYELDIAMDTATNNGISVGRSPDGRTHNIGKYGA 430
 Db 715 AS-----DRRQADRENLLKGSSVGYEIEAEIIPAQSNVSEFGFRLREGAGKRTVWGKT 771
 QY 431 ---DLVYDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILDVDTQSVVEFVNAGHTVLSQ 487
 Db 772 KENEIYDRSLSGDTGFSERFTTHQAPLQPDNRVRKLRIFVDSSLEVFEGDGRVVFSE 831
 QY 488 OVHFAEGDGTISLYTDGGPAHFTGIVVREI 517
 Db 832 VIFPDPAHREMSLFTVGGEVNWSLKVHAL 861

RESULT 7

AAV91928

ID AAV91928 standard; Protein: 1277 AA.

XX AC AAV91928;

XX AC AAV91928;

DT 19-JUL-2000 (first entry)

XX PAenibacillus pabuli 2,6-beta-D-fructan hydrolase.

DE 2,6-beta-D-fructan hydrolase; anticarie; neuroprotective; biofilm;

KW hydrolyze slime; detergent; fructose production; sweetener.

XX PAenibacillus pabuli.

OS PAenibacillus pabuli.

XX Key Location/Qualifiers

FH Protein 25-1277

FT /label= mature_protein

FT /label= mature_protein

XX

PN WO200017331-A1.

XX 30-MAR-2000.

XX 17-SEP-1999; 99WO-DK00495.

XX 18-SEP-1998; 98DK-0001173.

XX 09-DEC-1998; 98DK-0001623.

XX (NOVO) NOVO-NORDISK AS.

XX Moller S, Johansen C, Schaefer T, Ostergaard PR, Hoeck LH;

XX WPI; 2000-283564/24.

XX N-PSDB; AAA08533.

XX New polypeptides with 2,6-beta-D-fructan hydrolase activity, useful

XX e.g. for degrading microbial biofilm, hydrolyzing slime, production of

XX fructose oligomers and treatment of dental plaque

XX Claim 1; Page 111-115; 123pp; English.

XX AAY91927-29 are polypeptides from Paenibacillus amylolyticus, P. pabuli

XX and P. macerans, respectively, which have 2,6-beta-D-fructan hydrolase

XX activity. These are preferred enzymes of the invention. The enzymes

XX can be used to remove microbial biofilm from surfaces, particularly for

XX reduction of biocorrosion, for treatment of dental plaque and for

XX treatment of multiple sclerosis (eliminating biofilm from the lungs; to

XX hydrolyze slime; in detergents and for production of fructose oligomers

XX from levan or phleum (useful as sweeteners or for conversion to

XX fructose). The nucleic acids that encode the enzymes are useful for

XX recombinant production and as source of probes and primers for

XX identification and cloning of related fructan hydrolases (all claimed).

XX SQ Sequence 1277 AA;

Query Match 16.5%; Score 468.5; DB 21; Length 1277;

Best Local Similarity 29.1%; Pred. No. 8.4e-32;

Matches 169; Conservative 74; Mismatches 243; Indels 93; Gaps 25;

QY 4 AISRAVLOGAGA-----GALALIFGAVPAA-----RASAPG----- 37

Db 320 ASSNRKVPQSGARHHIEVIASGLIQVYVDGYTPAAVEVTDKSYAKNAGLVVOQGWAY 379

QY 38 -----SLRAVYHMTPPSGWLCDDPQRYTHGAYQLYLHSDQNGPGW 82

Db 380 FQDIYMTESMYKENYRPOYHSPRLGSASDPNGLVYEGEYHLF--HOD---GGTWA 433

QY 83 HASTTDGVAFTTHGTVMPLRDPFPVWGSAGVGTANTAGF-----GAGAVVALATQPTDG 137

Db 434 HAVSSDLINWKRLPIALPWNDOGHVWGSALADLNNASGLFTDGGKGLIAYTTSYHPDK 493

QY 138 VRKYOEQVLYWSTDGGTFT-ALPDPPIVNTDGRAATTPAEIENAEW-FRDPKIHMDTAR 195

Db 494 PGNORIGLAYSTDGGRNMOYAKERPVIDPNKNGDDP-----GSWDFRDPKVVROEDH 548

QY 196 GEWCVIGRLRYAAYTSPNLRDWTLRNFDPYHNLGGI-ECDFLFEITAD-DGTRHW 253

Db 549 NRWVWVSGGDHIFRFTSTNLLDWTLDNFGYDGVGVGWECPDLQLPVDCTGQRKWV 608

QY 254 LAASNDAYGIGLPMTYA-----YWTG--TWGGEQFADDLTPQWL--DWGWDWYAAVWTWS 305

Db 609 LLISTGAN---PKTQGSDAEYFVGLQATADG-KFLNDHPAGQVRLTDYKREYASMSFAN 663

QY 306 IDAPETKRLATAMNNKYAARDVPTDASDYGNGONSIVRELRLARQPGGWYTLTSPVA 365

Db 664 M--PNQRKVMIAWMTNWDYFP-EFTSS---WKGLTIPREVSURTDEG-VRLVQVPT 716

QY 366 ALTNVVTATTLTPRTVDGSVAVLPWNG---RAYEIEDI-----AMDATNVGISVGRSPDG 419

Db 717 ELQKLRLHLYSAQQWTVGPKSKNPLEGLTAGAYEIAEVEIPEANSSVTEFQFL-ROREG 775

QY 420 TRHTNICKYGAD---LYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILDVDTQSV 476


```
XX SQ Sequence 516 AA;
Query Match 15.1%; Score 427; DB 21; Length 516;
Best Local Similarity 27.2%; Pred. No. 9.9e-29;
Matches 145; Conservative 92; Mismatches 221; Indels 76; Gaps 21;

Oy 24 GGAVPAAARASAGSLRAYHYHTPPSGWLCDPQRPVTTTHGAYQLYLHSDQNNPGG--W 81
Db 16 GLTLPSQAQSN---DYRSPHYTPDOYWMNEPGLIKIGTWHLFFQHNPNTANWGNICW 72

Oy 82 DHASTTGDVATHTTCTVPLRPDPFVWSGSAVVGTAANTAGGAGA---VVALATOPTDGV 138
Db 73 GHATSTDLHMAHKPTATADENGVEAFTGTATYDPNNSTGLGDSANPPYLAFTGYTTS- 131

Oy 139 RYQOQYLYWSDGFTFTALPDVIVNTDGRAATTPAEIENAEFRDPKIHWDTARGEW 198
Db 132 SQTODORLAFSDNGATWTKFQGNPIISTSQEA---PHDITGGLESRDPKVFHFRQSNW 188

Oy 199 VCVI---GRLRYAAFTSPNLRDWTLRRNFDYPNHALGGI-----ECPDLFEITADGT 249
Db 189 IMVLAHGQDKLSFWTSADTINWTWQS---DLKSTSLNGLSSDITGWEPDMFELPV-EGT 245

Oy 250 RH--WVL-----AASMDAYIGLPMYAYWTGTWGEQFHAD--DLTPOWLWDGWDWYAAV 301
Db 246 BETTWVMVMTAEGSPAGGNGVLAI---TGSFDCKSFTADPVDASTWMLDNGRDFDGL 301

Oy 302 TWPSIDAPETKRLATAMNNKYYAARDVPTDASDGYNGSONSVIRELRLAROPGGWYTLIS 361
Db 302 SWNVNPSADGRRIIAAVNNSY---GSPNPTTT---WKMLSPFRTLSL-KKVGTOOHQVQ 354

Oy 362 TPVAALTNVYVATTPLPRTVDGSAVLPWNGRAYEIELDIADW--DTATNVGISVGRSPDG 419
Db 355 QPITELDTISTSLQILANOTITPGQTLSSIRGTALDVRVAFYPDAGSVLSLVRK---G 411

Oy 420 TRHNTICKY---GADLYVDRGPSDLAGYSLAPYSRAAPIDPGARSVH-----464
Db 412 ASEQTVIKYTSDFATLSVDRTESGDISY-----DPAAGGVHTAKLEEDGTGLV 459

Oy 465 -LRILVDTSQSEVVFVFNAGTIVLSQVHFAEGDTGISLTDGGAHFTGVVREI 517
Db 460 SIRVLVDTCSEVEFGGQGEAVISDLIFPSSDGLALEVTCGNAVLSQVDVRSV 513

RESULT 10
AAW44863
ID AAW44863 standard; Protein; 533 AA.
XX AC AAW44863;
XX DT 29-SEP-1998 (first entry)
XX DE C. utilis INV1 protein.
XX KW URA3; HIS3; INV1; orotidine 5'-monophosphate decarboxylase; enzyme;
XX KW biosynthesis; ribonucleotide; uracil; histidine; auxotroph; invertase;
XX KW imidazole-glycerol-phosphate dehydratase; beta-fructofuranosidase;
XX KW genetic marker; plasmid; transformation.
XX OS Candida utilis.
XX PN WO9814600-A1.
XX PD 09-APR-1998.
XX PF 03-OCT-1997; 97WO-CU000005.
XX PR 03-OCT-1996; 96CU-0000082.
XX PA (INGG-) CENT ING GENETICA & BIOTECNOLOGIA.
XX PI Besabe Tuero L, Chavez Espinoza FP, Delgado Boada JM;
XX PI Gonzalez Martinez ME, Paifer Reyes E, Rivero Baeza T;
```

```
PI Rodriguez Menocal L;
XX WPI: 1998-240099/21.
DR N-PSDB; AAV19464.
XX
PT New Candida utilis auxotrophic mutants and DNA markers - useful in
PT transformation systems to allow selection of transformants, e.g. for
PT heterologous protein production
XX
PS Disclosure; Page 29-31; 63pp; Spanish.
XX
CC This is the amino acid sequence of the Candida utilis INV1 enzyme
CC invertase (beta-fructofuranosidase). The invention relates to novel
CC C. utilis strains lacking one of the URA3 (AAV19462), HIS3 (AAV19463), or
CC INV1 genes resulting in auxotrophic mutants. The genes can be used
CC as markers in plasmids for transforming C. utilis. This organism has
CC not been used as a host for transformation due to difficulties in
CC obtaining reliable markers. Straightforward selection of transformants
CC using the new auxotrophic mutants as host cells is now possible.
XX
SQ Sequence 533 AA;
Query Match 14.7%; Score 416.5; DB 19; Length 533;
Best Local Similarity 25.9%; Pred. No. 8.7e-28;
Matches 132; Conservative 87; Mismatches 182; Indels 109; Gaps 22;

Oy 40 RAVYHMTTPSGWLCDPQRPV---TTHGAYQLYLHSDQNN---NGPGWDHASTTDCGVAFTH 94
Db 30 RPLVHLTPPVGMNDPNGLFYDSSSTHYVYQYNNPNTINGLPLYWGHATSDLLTWDH 89

Oy 95 HG-TVMPLRPDPVWSSGSAVVGTAANTAGGAGA-----VVALATOPTDGVRYQEQVLY 147
Db 90 HAPALGPENDEGIVSGSIVDYNSTGFFDSTREQRIVAIYTNLPLDV---ETQDIA 146

Oy 148 WSDGCGFTTALPDVIVNTDGRAATTPAEIENAEFRDPKIHWDTARGEWVCVIGRLR- 206
Db 147 YSTDGGYTFEKYENNPVIDV-----NSTQFRDPKVIWYEETEOWMVTAKSQE 194

Oy 207 -YAAFTYTPNLRDWTLRRNFDYPNHALGIECPDLFEITAD-----DGRHVLAAASMDA 260
Db 195 YKIQIYTDNLKDWLSASFSTKGVGVGYECPLFEATENPKSGDPEKKVMVVLAINP 254

Oy 261 YGIGLPM---TYAVWTGTWGEQFHADDLTPOWLWDGWDWYAAVTPWPSIDAPETKRLAIA 317
Db 255 ---GSPLGGSINEXEYVGDFNGTEFIPDDDDATRFMDTKDFYAFQAF--FNAPENRSIGVA 309

Oy 318 WMNNKYYAARDVPTDASDGYNGSONSVIRE-----LRLAROPGGWYTLSTTP 363
Db 310 WSSNNQY-SNQVPD--PDGYRSSMSSIREYTLRYVYSTNPESQQLILCORP---FEVNETD 363

Oy 364 VAALTNVYVATTPLPRTVD---GSAVLPWNGRAYEIELDIADWDTATNVGISVGRSPDGT 420
Db 364 LKVVEEYKVSNSL---TVDHFTGSSFANSNTTGL---LDF-----NMTFTVNGTTDVT 411

Oy 421 RHTNIGKYGADLYVDRGPSDLA---GYSL-----APYSR 451
Db 412 QKDSV---TFELRIKSNQSDAIALGYDYNNEQFYINRATESYFQRTNQFOERSTWTVQ 468

Oy 452 AAAPIDPCARSVHLRIILVDTQSVEVFVNAAG 481
Db 469 PLTITESGDKQVLYGLVDNINILELYFNDG 498

RESULT 11
AAW72566
ID AAW72566 standard; Protein; 490 AA.
XX AC AAW72566;
XX DT 23-NOV-1995 (first entry)
XX DE Penicillium purpurogenum endo-inulinase.
XX
```

KW Penicillium purpurogenum variety rubrisclerotium; endo-inulinase;
 KW inulo-oligosaccharides; inulin.
 XX
 OS Penicillium purpurogenum.
 XX JP07059574-A.
 PN
 XX 07-MAR-1995.
 PD
 XX 23-AUG-1993; 93JP-0229448.
 XX
 XX 23-AUG-1993; 93JP-0229448.
 PR
 XX (MITK) MITSUI TOATSU CHEM INC.
 PA
 XX WPI; 1995-135898/18.
 DR
 DR N-PSDB; AAQ86692.
 XX
 XX A new DNA encoding endo-inulinase - useful for the production of
 PT inulo-oligosaccharides
 XX
 PS Claim 1; Pages 4-5; 8pp; Japanese.
 XX
 XX AAQ86692 encodes AAR72566 the Penicillium purpurogenum variety
 CC rubrisclerotium endo-inulinase. The endo-inulinase can be used
 CC to prepare an inulo-oligosaccharide from inulin.
 XX
 XX Sequence 490 AA;

Query Match 13.7%; Score 389.5; DB 16; Length 490;
 Best Local Similarity 26.0%; Pred. No. 1.8e-25;
 Matches 133; Conservative 87; Mismatches 225; Indels 67; Gaps 19;

QY 40 RAVYHMTTPSGWMLCDPQRPVTHGAYQLYVLSHQNN--GPGGDHASTTGDVAFTHHGT 97
 Db 4 RTEHFCPAENWNEPGLIKIDSTWHLEYQADPTANVWNECHGATSSDLLHWDHLPV 63
 QY 98 VMLRDPFPVWGSVAVGTANTAGFAGA---VVALATOPTDGVKRYQEQYLYWSTDGFG 154
 Db 64 AIPVENGIESFTGTSYDANNTSSLTSTNPPYLPFTGTYTSS-NGTQDRLAYSTDLCG 122
 QY 155 TETALP-DPVIIVTDCGRAATPAEIAENAEFRDPKIHWDTARGEWCVI--GRLRYAAY 211
 Db 123 TWLKFSGNFI---SAALEAPHVDV7GGLESKDPKVFHEPBGKWMVLAHGGQKLTFW 178
 QY 212 TSNLRDMLTRNF-----DYPNHALGIECPDLFEITADGTRH--WVL----AASMD 259
 Db 179 TSLDAKSWTMSDLLASQIEGPPS-SVTCWEVPDMFQLPI-QGTNETTWIIFTPAQGSP 236
 QY 260 AYGIGLPMYAYWTGTWDEQFHAD--DLTPQWLDGWWDYAAVTPWPSIDAPETRLATA 317
 Db 237 AGNGV---VAITGSGFDGETFLANPVDSTLWLDYGRDFDGAMSWENVPASDGRLLIAA 292
 QY 318 WMNNWYAAADVPDASDCYNQNSIVRELRLARQPGWYTLTSTPVAALTNVYVATTTL 377
 Db 293 VMNSY---CSNPPTNT---WKMLSPPTLTLEKTSKQY-FLOQPIABELSTVDNALASI 345
 QY 378 PORTVDGSVLPWPNRAYEIELDIADWTATNVIISVGRSPDTRHTNI--GRYGADLYVD 435
 Db 346 QNOTIAPKOTLLSSIHGSSLDVRIAFSVDSGATLSLAVRKGGSEQTVIRYSOSNSTLSVD 405
 QY 436 RPSDLAGYSLAPYSRAAPIDQARSVH-----LRILVDTQSVVEFVNAGH 482
 Db 406 RTASGDIST-----DPAAGGIHSAQLARDNTELVYLRVLVDTCSEVEFGGQGE 453
 QY 483 TVLSQOVHFAEGDGTGISLTYDGPAPHTGIVV 514
 Db 454 AVISDLIFPSNSSLGSLSEVIGGTATLQSV 485

RESULT 12
 AAY25346
 ID AAY25346 standard; Protein; 750 AA.

XX AAY25346;
 AC
 XX 03-SEP-1999 (first entry)
 DT
 XX Bacillus sp. L7 endo-levanase protein.
 DE
 XX Endo-levanase; L7; slime; control; industrial water; pulp; paper;
 KW treatment; levan removal; industry.
 KW
 XX Bacillus sp.
 OS
 XX WO9931020-A1.
 PN
 XX 24-JUN-1999.
 PD
 XX 16-DEC-1998; 98WO-FI00987.
 PF
 XX 12-JUN-1998; 98US-0089045.
 PR
 PR 16-DEC-1997; 97US-0069801.
 XX
 XX (CULT-) CULTOR CORP.
 PA
 XX Miasnikov A;
 XX
 XX WPI; 1999-405013/34.
 DR
 DR N-PSDB; AAX78782.
 XX
 XX Using an endo-levanase enzyme preparation
 PT
 XX Disclosure; Page 18-21; 31pp; English.
 PS
 XX

This invention describes a novel method for the treatment of industrial
 CC waters with an endo-levanase enzyme preparation which results in a
 CC dramatically improved levan removal compared with current methods. The
 CC method reduces the amount of slime present in industrial waters and
 CC comprises contacting industrial water containing at least one slime
 CC material with an endo-levanase enzyme preparation at a temperature and
 CC pH condition sufficient to maintain activity of the enzyme. The
 CC industrial waters treated are especially from pulp or paper processes.
 XX
 XX Sequence 750 AA;

Query Match 13.6%; Score 386; DB 20; Length 750;
 Best Local Similarity 31.9%; Pred. No. 6.9e-25;
 Matches 115; Conservative 44; Mismatches 144; Indels 58; Gaps 16;

QY 40 RAVYHMTTPSGWMLCDPQRPVTHGAYQLYVLSHQNNGPGGDHASTTGDVAFTHHGTVM 99
 Db 398 RQYHYSPIRGSASDPNGLVYFEGEYHLP--HQD---GGQNAHAVSRDLIHWKRLPIAL 451
 QY 100 PLRDPFPVWGSVAVGTANTAGF---GAGAVVALATOPTDGVKRYQEQYLYWSTDGFG 154
 Db 452 PNWDLGHVWGSASAVADTTNASGLFGSGGKGLIAYVTSYNPDHNGNQKIGLAYSTDGR 511
 QY 155 TETALPD-PVIIVTDCGRAATPAEIAENAEFRDPKIHWDTARGEWCVIIGRLRYAAY 212
 Db 512 TWKYSEEHFVVENPKTGEDP-----GGWDFRDPKVRDEANNRWVWVYSGGDHRLFT 566
 QY 213 SPNLRDWTLRRNFDYFNHALGGI-ECPLDLEITADDG-TRHWVLAASMDAYGIGLPMYTA 270
 Db 567 STNLNLTWLTDFGCGAYIRGGWECPLDLPVSGSKRKKVLMISTCAN----PNTQG 622
 QY 271 YWTGTWDEQFHADDTLPQW-----LDWGNDWYAAVTPWPSIDAPETRLATA 317
 Db 623 S-----DAEYF-IGDLITPEGKFINDPACTVLKTDWGKEYYASMSFS--DMPDGRRLMLA 674
 QY 318 WMNNWYAAADVPDASDCYNQNSIVRE-----LRLARQPGGWYTLTSTPVAAL 368
 Db 675 WMTNWDYFP-SFPT---TGWKGQLSIPROVSLKETETEGRMHOTPIEELAQRLSPVLHIT 730
 QY 369 N 369

CC the S. cerevisiae SUC2, GAL4 and GAL1 genes. The inulinase gene appears
CC to be regulated solely by glucose repression which allows the
CC construction of a strong, non-repressible promoter by exchanging the
CC MIG1 DNA binding site in the inulinase promoter. The regulatory region
CC of the inulinase gene comprises a promoter, an upstream activating or
CC repressing sequence (UAS or URS), a terminator as a regulatory region,
CC and a sequence encoding a secretory signal necessary for secreting a
CC gene product from a yeast. This regulatory region may be used to
CC prepare transformed yeasts for the production of desired gene products,
CC eg. protein or metabolites.

XX Sequence 556 AA;

Query Match 12.9%; Score 366; DB 15; Length 556;
Best Local Similarity 27.0%; Pred. No. 2.5e-23;
Matches 132; Conservative 72; Mismatches 181; Indels 104; Gaps 24;

QY 40 RAYVHTPPSGHLCDPQRPVTHGA-----YQLYLHSDN---NGPGGWDHASTTD 88
DB 39 RPSVHTPSHGWNNDP-----NLWDYDAKEEDHLYQYNPAATWGTPLYGHAVSKD 92
QY 89 GVAFTHHGTVM-PLRPDPFVWGSAAVGTANTAGF-----GAGAVVALATOPTDGV 138
DB 93 L7SWTDYGASLPGSDAGAFSGSMVIDYNTSGFNSVDPRQRAVAVWTLKGPS--- 149
QY 139 RYQOYLWSTDGGFTFTALPDPVIVNDGRAATTPAEIENAEFRDPKIHWDTAR--- 195
DB 150 ---QAQHSYSLDGGYTFQYSDNAVLDI-----NSSNFRDPKVFWEHGE 194
QY 196 -GEWVCVIGRLAY--AAFTYSPNLRDWTLRRNFDYPNHALGIECPDL---FEITADDG 248
DB 195 DGRWIMAEOSVFSVLFYSSPNLKNWLTESNFTHHGTGTQYCEPGLVKVPYDSVADSS 254
QY 249 TRH-----WVLAASMDAYG-IGLPMTYAYWTGTWGEQFHADDLTPQWLDGMDWYA 299
DB 255 SNSDSKPSANVLFVSNIPGGLGSSVT-QYEVGDFNGTHETPIDDQTRFLDMGKDYA 313
QY 300 AVTSPSIDAPEKRL-AIAMNNKYAARDVPTDASDGYNGQNSIVRELRL---ARQGG 355
DB 314 LQTF--FNTPNKDYGTAWASNWOY-AQAAPTDP---WRSSMSLVROFTLKDFTNPN 367
QY 356 WYTL-STEPVAALTNYVTATTLTPORTVDGSALVLPNGRAYEILDIAWDATNVGISVG 414
DB 368 ADVLNSQPV---LNYDALRKNGTYSINVTTSSENGK--KIKLD----- 408
QY 415 RSPDGRTHNIGYGADLYVDRGPSDLACYSLAPYSRAAAPIDPGARSVHLRLVDTQSV 474
DB 409 -NPSGSLFHL------YVENGSPDIKSNVFDLSLYFKGNDD--NEYRLGYETNGG 459
QY 475 EVFVNAGHT 483
DB 460 AFFLDRGHT 468

RESULT 15
AAP40020

ID AAP40020 standard; Protein; 532 AA.

XX

AC AAP40020;

XX 30-NOV-1991 (first entry)

XX Saccharomyces cerevisiae invertase.

XX Signal sequence; invertase; vector; leukocyte interferon.

XX Saccharomyces cerevisiae.

XX EP127304-A.

XX 05-DEC-1984.

XX 24-APR-1984; 84EP-0302722.

PF

XX

PR 25-APR-1983; 83US-0488337.
XX (GETH) GENENTECH INC.

XX Chang CN, Matteucci MD, Hitzeman RA;
XX WPI; 1984-301996/49.

DR N-PSDB; AAN40009.

XX Yeast expression vehicle - contains DNA sequence of yeast promoter
PT linked to DNA sequence encoding homologous protein.

XX Disclosure; Fig. 13; 53pp; English.

XX The DNA encoding the signal peptide of S. cerevisiae invertase may
CC be used in a new yeast expression vector for the expression of, esp.
CC human leukocyte interferon-alpha. The expression vector comprises a
CC yeast promoter operably linked to the signal sequence. This is attached
CC to the coding sequence of the protein. Yeasts transformed with this
CC vector produce and secrete protein normally heterologous to them.
CC The protein can be recovered from the medium free of presequences
CC etc. Human leukocyte interferon can be used as an antiviral or
CC antitumour agent. The promoter is esp. the invertase promoter or
CC the yeast 3-phosphoglycerate kinase promoter.

XX Sequence 532 AA;

Query Match 12.7%; Score 359.5; DB 5; Length 532;
Best Local Similarity 26.0%; Pred. No. 8.9e-23;
Matches 143; Conservative 81; Mismatches 228; Indels 99; Gaps 25;

QY 19 LALIFGAVPPAARASAGSLRAVYHMTPTSGWLCDP-----QRPVTHGAYQLYLHS 72
DB 7 LELLAGFAAKISASMTNETSDRPLVHETPKGMNDPGLWYDEKDAWH---LYFOYN 62
QY 73 DON---NGPGGHDHASTTDGVAFTTHG-TVMPLRDPFVWGSAAVGTANTAGFGAGAV- 127
DB 63 PNDTVMGTPLEFGHATSDDLTNWEDQPIAIAPIKNDGAFSGSMVVDYNTSGFFNDTID 122
QY 128 -----VALATQPTDGVKRYQOYLWSTDGGFTFTAL-PDPVIVNDGRAATTPAEIENA 181
DB 123 PRQRCVAINTYNTP---ESEEQYISYSLDGGYTFTEYQKNPVL-----ANS 166
QY 182 EWRDPKHWDTARGEWCVICGLRL--YAAFTYSPNLRDWTLRRNFDYPNHALGG--IEC 237
DB 167 TQFRDPKVFWEYPSQKWIIMTAASKQDYKIEIYSSDDLKSKWLESAP--ANEGLFGYQEC 224
QY 238 PDLFEITA--DDGTRHWVLAASMDAYGICLPMTYA---YWTGTWGEQFHADDLTPQWLD 292
DB 225 PGLIEVPTEQDPSKSYWVNFISINP---GAPAGGSFNQYFVGSFNGTHEAFDNGSRVVD 281
QY 293 WGDWYAAVTPSIDAPETKRLAIAAMNNKYAARDVPTDASDGYNGQNSIVRELRL--- 349
DB 282 FGKDYALOTFTDPTGYSALGIASWNEIYA-FVPTNP---WRSSMSLVKRFSLNTE 337
QY 350 ---AROPGGWYTLTTPVAALTN-----YVTATTLTPDRT--VDGSALVLPNGRAYEIEL 399
DB 338 YOANPETELINLKABPILNISNAGPWSRPATNTTLTKANSYND-----LSNSTGTLEFEL 393
QY 400 DIAWDATNVGISVG-----RSPDGRTHNIGYGADLYVDRGPSDLACYSLAPY 449
DB 394 VYAVNTQPTISKSVFADLSLWPKGLEDPPEYLRMGFEVSASSFFLDGRGSKVKFKENPY 453
QY 450 --SRAAAPIDPGARSVHLRI-----LVDTQSVFVFNAGHTVLSQOVHFAEGD----- 495
DB 454 FTNRMSVNNQPEKSENDSLYKVGILLDQNIILELFNDGDVYSTNTYFTTGNALGSVNM 513
QY 496 -TGIS--LYTD 503
DB 514 TTGVNDLFIYD 524

Search completed: May 6, 2003, 12:32:14
Job time : 36.9327 secs

THIS PAGE BLANK (USPTO)

Db 237 FEMTAGDGRHWVFGASMDAYSIGLPMTFAYWTGSWNGTAFIADNLTPQWLDPQWLDWGDWYAA 296

QY 301 VTWPSIDAPETKRLATAMNNKYYAARDVPTDASDGYNGONSIVRELRLARQPGGWYTL 360
Db 297 VTPVAEAPETKRLATAMNNKYYAARNVPTDASDGYNGONSITRELRLERQSGGWYTL 356
QY 361 STPVAALTNVYATTTLPDRTPDGSALVPWNGRAYEIELDIANDTATNVGSGSPDGT 420
Db 357 STVPALSNVYATSTTLPDRTPVNGSPVLPWNGRAYEIELEISWDTAAVNGVSGSDGS 416
QY 421 RHTNICKYGADLVDRGPDLAGYSLAPYSRAAAPIDPGARSVHLRLVDTQSVYEVFNA 480
Db 417 RHTNICKYGDELVDYDRASSQSYALAPYTRAAPIDANARSVHLRIEVDQTSQVEFVNS 476
QY 481 GHTVLSQVHFACDYGISLYTDGGAHFTGIVVREIG 518
Db 477 GHTVVSQVHFAGADYGISLYADGGPANFTGITIREFG 514

RESULT 2
E70035
levanase homolog yveB - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: E70035
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.: Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.: Ehrlich, S.D.; Emmerson, P.F.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim
wood, C.R.; Hensaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.
Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber
A:Authors: Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigu
K.; Ogilwa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwol
Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleid
A:Authors: Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serro
amakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol
Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa,
A:Authors: Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: E70035
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-516 <KUN>
A:Cross-references: GB:290121; GB:AL009126; NID:g2635827; PIDN:CAB15451.1; PID:g2635959
A:Experimental source: strain 168
C:Genetics:
A:Gene: yveB
C:Superfamily: Penicillium purporogenum inulinase

Query Match 33.3%; Score 944; DB 2; Length 516;
Best Local Similarity 41.8%; Pred. No. 4e-62;
Matches 205; Conservative 71; Mismatches 176; Indels 38; Gaps 15;

QY 40 RAVYHMTTPSGWLCDDPQRPVTHGAYQLYLLHS--DQNGPG-CWDHASTTGGVAFTHHGT 97
Db 44 RAYHFTTDDKNDPQRPVTHGAYQLYLLHS--DQNGPG-CWDHASTTGGVAFTHHGT 103
QY 98 VMP--LRPDPFVWMSGSAVGTANTAGFAGAVVALATQPTDGVKRYQEOYLWSTDGGT 155
Db 104 AIPKYNPDGDIWTSVVDKNTAGFAGAVVALATQPTDGVKRYQEOYLWSTDGGT 162
QY 156 F-----TALPDVIVTNGRAATTPAEIENAEWFRDPKHWDGTARGWCVIGRLRYAA 209
Db 163 FKPYSGNPVMPN-----GTDD-----FRDPKVIWDDQDNKVMVMAEGSKIG 205
QY 210 FYTSPNLRDWTLLRRNFDYPNHALGGTECPDLFEITADDGTRHVLAAASMDAYGIGLPMY 269
Db 206 FRESNLDKWHYTSGF-FPEQA-GWCEPDLYMNRASDGTNRKWLUGASGKNGPWPKNPT 263
QY 270 AYWGTGWDGEQFHADLTLPQWLDGWDYAAVTPSIDA--PETKRLAIAWNNKYYAAR 327
Db 264 AYWGTGSDCKEKFADQTEAQWLDYGDWYGGVTFEDSKSDTDPLEKRYALAWNNNDY-AN 322

RESULT 3

G97118

levanase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001

C:Accession: G97118

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L

J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: G97118

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-514 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK9738.1; PID:g15024743; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC1773

C:Superfamily: Penicillium purporogenum inulinase

Query Match 31.0%; Score 880.5; DB 2; Length 514;

Best Local Similarity 40.7%; Pred. No. 2e-57;

Matches 200; Conservative 76; Mismatches 179; Indels 37; Gaps 14;

QY 40 RAVYHMTTPSGWLCDDPQRPVTHGAYQLYLLHS--DQNGPG-CWDHASTTGGVAFTHHGT 97

Db 45 REVYHFTVPMKNDPQRPVTHGAYQLYLLHS--DQNGPG-CWDHASTTGGVAFTHHGT 104

QY 98 VMP--LRPDPFVWMSGSAVGTANTAGFAGAVVALATQPTDGVKRYQEOYLWSTDGGT 155

Db 105 CIPKYNENGDWISGFSVVDQAQNTAGFAGAVVALATQPTDGVKRYQEOYLWSTDGGT 163

QY 156 FTALPD-PVIVNTDGRAATTPAEIENAEWFRDPKHWDGTARGWCVIGRLRYAAYTSP 214

Db 164 FKPYSNQPIPN-----PCTKD-----FRDPKIIWDFKNNKVMVLAEGTKIGFVSY 211

QY 215 NLRDWTLRRNFDYPNHALGGTECPDLFEITADDGTRHVLAAASMDAYGIGLPMYAYWG 274

Db 212 NLKNW--QHTGDFFTNIGIVCEPDIFMNSDNGNTKWLGTSAKGVSEPTNYAYWG 269

QY 275 TWGDEQFHADLTLPQWLDGWDYAAVTPSIDAET--KRLAIAWNNKYYAARDVPTD 332

Db 270 NYDGKFIADISTPKWLDYGFWDYAAVTPSESNOSKRLKRYALAWNNNDY--ANNTP- 327

QY 333 ASDGYNGONSIVRELRLARQPGGWYTLSTPVAALTNVYATTTLPDRTPDGSVLPWNG 392

Db 328 IQNRFNGMSIVREITLSKQKNTYSLSKPKRIENITTSIDQFKQISVKGLKHLKVG 387

QY 393 RAYEELDTAWDTATNVGSGSPDGTTHNIGKY--GADLYVDRGPDLAGYSLAP-- 448

Db 388 KAYQDLDTDINWDAKNVGRILRESLDKKRHIDVIGIFTEGKYSVNR-----AYTNPDK 441

QY 449 ---YSRAAAPIDPGARSVHLRLVDTQSVYEVFVWNGTATVLSQVHFACDYGISLYTDGG 505

Db 442 SKKYVESRAPFDLNNKVVHLRFVDKVSVYEVFIDDKIYTSNEVFPKPKDGITLFSIKG 501

Qy	506	PAHFTGIVVREI	517
		:	
Db	502	KAVFKNITIKHI	513

RESULT 4

A53206
sucrose/fructan hydrolase (EC 3.2.1.-) - *Bacillus polymyxa*
C:Species: *Bacillus polymyxa*
C:Date: 05-May-1995 #sequence_revision 05-May-1995 #text_change 08-Oct-1999
C:Accession: A55206; S37558
R:Bezzate, S.; Steinmetz, M.; Aymerich, S.
J. Bacteriol. 176, 2177-2183, 1994
A:Title: Cloning, sequencing, and disruption of a levanasase gene of *Bacillus polymyxa* CP4
A:Reference number: A55206; MUID:94209215; PMID:8157587
A:Accession: A55206
A:Molecule type: DNA
A:Residues: 1-512 <BEZ>
A:Cross-references: EMBL:Z26651; NID:g406312; PIDN:CAA81392.1; PID:g406313
C:Genetics:
A:Gene: *leIA*
C:Superfamily: Penicillium purporogenun inulinase
C:Keywords: glycosidase; hydrolase

Query Match 19.8%; Score 562.5; DB 2; Length 512;
Best Local Similarity 31.2%; Pred. NO. 6.6e-34;
Matches 162; Conservative 75; Mismatches 213; Indels 69; Gaps 18;

Qy	40	RAYVHTPPSGWGLCDQRPVYTHGAYQLYLYLHSDQNGPG-----GWDHASTTGDGVAFTHH	95
Db	13	RPOFHYSPKNNWMDNGLYIFEGEYHLFYQHPTQDFGNNMHGHAVSXDLVHMTSL	72
Qy	96	GTVMRLRPDPVYMSGSAVGTANTAGF---GAGAVVALAT-----QPTDGRVKRYQEQY	145
Db	73	PPAIPGCEGAIFGSAVVDKNNTSGFDEEGSG-LVAIYTNEGNKSQP-----GRPOVOS	127
Qy	146	LYWSTGGFTFTALP-DPVYVNTDGRAATTPABIEAENWFRPKIHWDTAREWVCVIGR	204
Db	128	IAYSKDKGRWTYKEGNPVLPFTDTLD-----FRDPKVIWHDSESSMMINVLAV	175
Qy	205	LRVAAAFYTNLNRDWTLLRRNF---DYPNHALGGTECPDLEIFETADD--GTRHWVLAASM--	258
Db	176	RDRVEYFYNLAKENWSEFSGSDIPIHRIHGFECDFRIQVDEDLNTTKAILMLSVGD	235
Qy	259	-----DAYGIGLPMYAYWTGTWDGEOFHADDL-----TPQWLMDGWDNYAAVT	302
Db	236	RNGVNPDDPPPPAGGSGM-----MYFYSFGDKSFTDEALESIDTTKKWYDYSDFYAAYS	291
Qy	303	WPSIDAPETKRLAIAMNNHKYAAARDVPDASDGYNCONSIVRELRLARQPGWYTLTST	362
Db	292	WNGISNEDGRKILWGNWNNRYAT---TLPSKEWRGKTSIPRELOLRYPEG-LRLIOT	346
Qy	363	PVAALNYVYATTTLPDRTV-DGSAYVL-PWNGRAYEIELDIADWDTATNYGIVSGRSPDGT	420
Db	347	PINELSQLRPILSLQDLTIKPGMNYLSDISAAKABIIAEFETGTAVEFGFKVRKS--AN	404
Qy	421	RHTNIGK--YGADLYVDRGRPSDLAGSLAPYSRAAAPIDPGARSVHLRILVDTQSVFEV	478
Db	405	QETIIIGYNISNEELFVDRTKSSATDFHSDFTAIHKATMKPEHRIQLSYLDWSSVEFG	464
Qy	479	NAGHTVLVSQOVHFAEGDGTGISLYTDGGGPAHFTGIVVREI	517
Db	465	NHGKAIISDIFFPDFSGKGLYELVIAIGELRVWSLOLNDL	503

RESULT 5

C:Species: *Bacillus subtilis*
 C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 21-Jul-2000
 C:Accession: A27286; S06353; S11402; A69703
 R: Schoengendorfer, K.; Schwab, H.; Lafferty, R.M.
 levansase (EC 3.2.1.65) precursor - *Bacillus subtilis*
 A27286

Db 258 MQYFVGDFDTHPKNPNKVLWTDYGRFYAAVSWSDIPSDSRRLWLGWNSNQY-A 316
QY 327 RQYPTDASQYNGQNSIVRELRLARPGGWYTLTLLSPVAALTYVATTTLPDRTVDGSA 386
Db 317 NDVPTSP---WRSATSIIPRELKKAFTG-VRVQTPVKELETIRGTSKKWKNLTISPAS 372
QY 387 ---VLPWNGRAYEI--ELDIANDTATNVGISVGRSPDGRHRTNIGKYGADLYVDGRPSDL 441
Db 373 HNYLAGQSGDAYEINAEFKYSPGSAAEFGKVRTGENQETKVGYDRNRKALFVDRSESGN 432
QY 442 AGYSLA-PYSRAAPDPCGARSVHLRLVDTQSVFVFNAGHTVLSQQVHFAEGDTGISL 500
Db 433 DTNPAPNTGKETAPLKPVNGKVKLRIFVDRSVSEVFGNDGKQVITDIILPDRSSKGLEL 492
QY 501 YTDGGAHETGVVIREI 517
Db 493 YAANG-----GVKXSL 504

RESULT 6
A36915
fructanase - Bacteroides fragilis
C:Species: Bacteroides fragilis
C>Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 17-Mar-2000
C:Accession: A36915
R:Blatch, G.L.; Woods, D.R.
J. Bacteriol. 175, 3058-3066, 1993
A:Title: Molecular characterization of a fructanase produced by Bacteroides fragilis BF-1
A:Reference number: A36915; MUID:93259952; PMID:8491724
A:Contents: BF-1
A:Accession: A36915
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-622 <BLA>
A:Cross-references: GB:M83774; NID:q143970; PIDN:AAA22924.1; PID:q143972
A:Note: sequence extracted from NCBI backbone (NCBIN:131947, NCBI:P:131949)
C:Superfamily: beta-fructofuranosidase

Query Match 18.2%; Score 515; DB 2; Length 622;
Best Local Similarity 30.2%; Pred. No. 2.8e-30;
Matches 156; Conservative 77; Mismatches 195; Indels 88; Gaps 20;

QY 40 RAVYHMTPPSGWLCDPQRPVTHGAYQLYLHSDQNN--GPGGWDHASTDGTGVAFTHHGT 97
Db 133 RPLVHTPLVGMNDANGLVYKDGVEHLYFYQNPYSGMNGMHWGSHSVSKDLVHWEH--- 189
QY 98 VMLRPDPEP-----VWSGSAVVGTAAGAGAVVALATOPTDGVRYQBOYLWSTD 151
Db 190 ---LEPALARDTLGHIFSGSSVVDDANTAGYGAGAIYAFYTSASD--KNGQIQCMAYSTD 244
QY 152 GGTFTTAL-PDPVIVMTDGRAATPAEINAEWFRDPKIHWDTARGEWVCVIGRLRYAAF 210
Db 245 NGRTFTKYEKNPVLTDFG-----LKDFRPKVFWAPDQKWVWVSADKEMRF 293
QY 211 YTSPLNRLDWTLRNFD-----TPNHALGIECPDLFEITADGTRH--WVLAASMDAYGI 263
Db 294 YSSENLEKWTYASGMGEGYGVQPSQ----FECPDWELPVDGNPDHKKWALIVNVN---- 345
QY 264 GLPMTY-----AYWTGTWDGQFHADDL--TPQWLDGWDWYAAVTWPSIDAPETKRLA 315
Db 346 ---PGCYFGGSATQYFIDGDFGKFCVCDNRPETVKWLDGKDHATVCFNSNTG---DRTIA 400
QY 316 IAWNNWYAADVPDTPDASDNGQNSIVRELRLARPGGWYTLTLLSPVAALTYVATTT 375
Db 401 VPMNSNQY-ANVTPTRQ---FRSANALPRELSLYTQDGDYI-MAAAPVEETKSLRKESR. 455
QY 376 TLPDRTVDGS-----AVLPWNGRAYEIELDIANDTATNVGISVGRSPDGRHRTNIGKYAD 431
Db 456 EYPAFEVGDAYHVDLSLSDNKGAYETELELAAGSAEIMGLKLFNEKGENVDYIYISLPEKK 515
QY 432 LVVDGRGPSDLA--GYSLAPYSRAA-----APIDPGARSVHLRIIV 469

Db 516 LYMDRTKSGIVDFGKDSAPHAIEAHRDKONKSNINYVDDFALGTWAPVQK-AGNYKLDIFV 574
QY 470 DTQSVFVFNAGHTVLSQQVHFAEGDTGISLYTDGG 505
Db 575 DKCSVEIFLNGGKIAMTNLIFFTPPYNQMSFYSRGG 610

RESULT 7
JC4081
sucrase/fructanase precursor - Actinomyces naeslundii (strain T14V)
C:Species: Actinomyces naeslundii
C>Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 15-Oct-1999
C:Accession: JC4081
R:Norman, J.M.; Bunney, K.L.; Giffard, P.M.
Gene 152, 93-98, 1995
A:Title: Characterization of levJ, a sucrase/fructanase-encoding gene from Actinomyce
A:Reference number: JC4081; MUID:95129923; PMID:7828936
A:Accession: JC4081
A:Molecule type: DNA
A:Residues: 1-943 <NOR>
A:Cross-references: GB:U12274; NID:g515689; PIDN:AAA67876.1; PID:g515690
A:Note: The authors translated the codon ATG for residue 467 as Asn, TAC for residue
C:Genetics:
A:Gene: levJ
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-943/Product: sucrase/fructanase #status predicted <MAT>

Query Match 16.5%; Score 467.5; DB 2; Length 943;
Best Local Similarity 21.5%; Pred. No. 1.6e-26;
Matches 187; Conservative 83; Mismatches 217; Indels 381; Gaps 27;

QY 4 AISRRAVLQAGAGALALIFGG----AVPPAARASAPGS-LRAYVHMTPPSGWLCDPQRP 58
Db 19 AASTLASMPAAASGQARPVSAAAPNAPAPAPKADQGTGERWRPQSHYTPQKNMNDPGL 78
QY 59 VTHGAYQLYLHSDQNN--GPGGWDHASTDGTGVAFTHHGTVMPLRDPFVWSGSAVGT 116
Db 79 VYDGEYHMFYQNPYSGMNGMHWGSHSVSKDLVHWEHLYFYQNPYSGVFSGSAVIDT 138
QY 117 ANTAGFCA---GAVVALATOPTDGVRYQBOYLWSTDGGETFTAL--PDPVIVNTDGRA 171
Db 139 KNTSGLGSPDNPAWAVWTRADVGN--QSQSLAYSTDGKTWNLYNNGDPVL----- 189
QY 172 ATPPAEINAEWFRDPKIHWDTARGEWVCVIGRL--RYAAFTSPNLRDWTLRNFDYPN 229
Db 190 -----DIGSNE-FRDPKVFDQASGRWTVVWVSHATEHRVSYFSYSPDLIHWTEQSFEGEG 243
QY 230 HALGGIECPDLFEITADGTRH--WVLA-----255
Db 244 ITSAMWACPDFPLPVDGSSQEVKWLVTVTADSAYQFVGSWMDGTTFTTPEIPHYSGEG 303
QY 256 -----ASMDAYG-----262
Db 304 TTLADPENGYAGWKADGAAGFGSPATGDLPGHQKAYVDSFGSDADTGTLSDEFTVSS 363
QY 263 -----262
Db 364 SYVNLRTAGKHPNPQATGDMGGRLLAGFDGSGWEGTVEGSFAFAATPQCATPAQQL 423
QY 263 -----IGLPWTY-----AY-----271
Db 424 VNHSSAGLLNTYLDAAATGGSAPGTATSPFTTIDSAVLNLLMGGGNPRPEGADGGS 483
QY 272 -----W-----TGTW-----276
Db 484 RYSVVELIYDGKVVRSATGRNLEELNQSWDYSDLKGSQAIWVTTDTATGGGHILLDEV 543
QY 277 -----276
Db 544 RASDKKASPIADNTSYNLYVDGVKVVASATGNNSGTLWTSMNVAAKGRKARLVIEDRNG 603
QY 277 -----DGQEFHADDLTPOWLDGWDWYAAVTWPSIDAPETKRLATAWN 320

Db 604 NAEQWGLHMDQIIQSDTKARSGADVPR-LDYGKDYAAVTWQNV--PNCKRYQVGMWS 660
Qy 321 NWKYAARDVPTDASGYNGQNSIVRELRLARPGGWYILLSTPVAALTNVTA-----T 374
Db 661 NWAY-VRDLPITT---WRTANSTVREMGLTR-VNGKRLRTAQPVTALESRLTGQELIRKD 715
Qy 375 TTLP-DRVTGSAVLPMWNGRAYEIELDIAMDTATNVGTSVGRSPDGTHTNIG--KYGAD 431
Db 716 TDIPVGETSLGKAA---QGTSLDISVDSLSPASSPAGLKV--LDNGEQYTLIGYDSQAKQ 770
Qy 432 LYVDRGPSDLAGYSLAPYSRAAAPIDPCAR-SVHLRIILVDTSQVEVFVFNAGHTVLSQQVH 490
Db 771 LVDRTHSGVDFSPKFPARSTAPUSPSKGVHLRIILVDAHSVEVFAADGTPVITQTVY 830
Qy 491 FAEGDTGSLYTDGGAHFTGIVVREIG 518
Db 831 PQDQATGVSVAEGGTAHLGSLSLWHLG 858

RESULT 8
JE0301
inulinase (EC 3.2.1.7) - Aspergillus niger
C:Species: Aspergillus niger
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 07-May-1999
C:Accession: JE0301
R:Ohta, K.; Akimoto, H.; Matsuda, S.; Toshimitsu, D.; Nakamura, T.
BioSci. Biotechnol. Biochem. 62, 1731-1738, 1998
A:Title: Molecular cloning and sequence analysis of two endoinulinase genes from Aspergillus niger
A:Reference number: JE0301; MUID:95022189; PMID:9805373
A:Accession: JE0301
A:Molecule type: mRNA
A:Residues: 1-516 <OHT>
A:Cross-references: DDBJ:AB012771
C:Superfamily: Penicillium purporogenum inulinase
C:Keywords: glycosidase; hydrolase

Query Match 15.38; Score 434; DB 2; Length 516;
Best Local Similarity 27.18; Pred. No. 2.le-24;
Matches 144; Conservative 92; Mismatches 225; Indels 70; Gaps 19;

Qy 24 GGAVPPAARASAPGSLRAYHMTPPSGWGLCDPQRPVTHGAYQLYLSHDQNNPGG--W 81
Db 16 GLMPSQAQSN---DYRSYHFTPDQYWNNEPGLIKGTWHFFQHNPTANVWGNICW 72
Qy 82 DHASTTGVAFTHGTVMPLRPDPFVWNSGSAVGTANTAGFAGA---VVALATQPTDGV 138
Db 73 GHATSTDLHWAHKAPTAIDENGVEAFTGTAYDPNNTSGLGDSANPPYLAFTGYTTS- 131
Qy 139 RYQEQYLYWSTDGFTFALPDVIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGEW 198
Db 132 SQTQDQLAFSVDRNGATWTKFQGNPIITSOEA---PHDITGGLSREDPKVFFHRQSGNW 188
Qy 199 VCVI--GRLRYAAFTSPNLRDWTLLRRNFDYPNHALGII-----ECPDLFEITADG 249
Db 189 IWLVAHGQDKLSFTWTSADTINWTQW---DLKSTISGLSSDITGWEVPMDFELPV-EGT 245
Qy 250 RH--WVL---AASMDAYGIGLPMYAYWTGTWQEQFHAD---DLTPQWLDWGDWYAAV 301
Db 246 EETTVVWMTFAEGSPAGNGVLAI---TGSFDCKSFTADPVDASTMWDNGRDFDCA 301
Qy 302 TWPSIDAPETKRLAIAWNNKYYAARDVPTDASGYNGQNSIVRELRLARPGGWYLLS 361
Db 302 SWAVNPASDGRRIIAAVNNSV---GSPNPTTT---WKGLMSPFRTLSL-KRVGTQQHFVQ 354
Qy 362 TPVAALTYVTATTLTPRTVDGSAVLPMWNGRAYEIELDIAM--DTATNVGISVCRSPDG 419
Db 355 QPITELDTISTLSLOTLANQITPGOTLSSIRGTALDVRVAFYDPDAGSVLSLAVRKGASE 414
Qy 420 TRHTNIGKYADLVDRGPSDLAGYSLAPYSRAAAPIDPGARSVH-----LR 466
Db 415 QTVINYTQSNATLSVDRTESGDISY-----DPAAGGVHTAKLEDDGTGLVSIR 462
Qy 467 ILVDTSQVEVFVFNAGHTVLSQQVHFAEGDGTGISLTDGGAHFTGIVVREI 517

Db 463 VLVDTCSEVEFGQGEAVISDLIFFSDSDGLAEVTTGGNAVLSQSDVRSV 513

RESULT 9
JC7533
inulinase (EC 3.2.1.7) - Penicillium sp. (strain TN-88)
N:Alternate names: 2,1-beta-D-fructan fructanohydrolase
C:Species: Penicillium sp. (strain TN-88)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: JC7533
R:Akimoto, H.; Kiyota, N.; Kushima, T.; Nakamura, T.; Ohta, K.
BioSci. Biotechnol. Biochem. 64, 2328-2335, 2000
A:Title: Molecular cloning and sequence analysis of an endoinulinase gene from Penicillium sp. (strain TN-88)
A:Reference number: JC7533; MUID:21036892; PMID:11193399
A:Accession: JC7533
A:Molecule type: mRNA
A:Residues: 1-515 <AKI>
A:Cross-references: DDBJ:AB041337
A:Experimental source: strain TN-88, wild type
C:Comment: This enzyme, an acidic glycoprotein and an endo-acting inulinase as well, is
ase as the main products.
C:Genetics:
A:Gene: inuU
C:Superfamily: Penicillium purporogenum inulinase
C:Keywords: glycoprotein; glycosidase; hydrolase

Query Match 15.04; Score 425; DB 2; Length 515;
Best Local Similarity 27.68; Pred. No. 9.9e-24;
Matches 142; Conservative 92; Mismatches 232; Indels 48; Gaps 21;

Qy 31 ARASAGSLRAYHMTPPSGWGLCDPQRPVTHGAYQLYLSHDQNN--GPGGDHASTTD 88
Db 21 ARAVA-DDYRPAHFPCPAENWMNEPGLIQINSTWHLFYQADPAANVWNGECWGHATSSD 79
Qy 89 GVAFTHHGTVMPLRPDPFVWNSGSAVGTANTAGFAGA---VVALATQPTDGVRYQEQY 145
Db 80 LLHWDHLVPAIVPENGIESFTGTSYSDNNTSGLTSTNPPYLAFTGYTES-NKTQDOR 138
Qy 146 LYWSTDGSGFTTALP-DPVIWNTDGRAATTPAEIENAEWFRDPKIHWDTARGEWCVI-- 202
Db 139 LAYSTDLSQGTWVKFAGNPII---GAAQEAQDISGGLSREDPKVFFHAPSCKWVMVLAH 194
Qy 203 GRLRYAAFTSPNLRDWTLLRRNFD-----YPNHALGIECDLFEITADGTRH--WVL 254
Db 195 GGQDKLTWTSLDKAKNWTWVSDLSSSSQIEGFPSS-SITGWEVPMDFQLPI-QCIKKTWVL 252
Qy 255 ----AASMDAYGIGLPMYAYWTGTWQEQFHADDLTPQ--WLDWGDWYAAVTPSIDA 308
Db 253 IFTPAQGSPPGNGV----VALTGSFDGETFVADPVDPSTLWLDYGRDFDGLSALWENVPA 308
Qy 309 PETKRLAIAWNNKYYAARDVPTDASGYNGQNSIVRELRLARPGGWYLLSTPVAALT 368
Db 309 SDGRILIAAVNNSV---GSPNPTTT---WKGLMSPFRTAL-KQIGSKQYFLQQPAVELS 361
Qy 369 NYVTATTLTPRTVDGSAVL--PMWNGRAYEIELDIAMDTATNVGTSVGRSPDGTHTNIG 426
Db 362 TIDGSLTSIQNTITPNTQTLSSIHGTSLDIRMAFVIDSGATLSLAVRKG--GSEQTIVR 419
Qy 427 KY--GADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRIILVDTSQVEVFVFNAGHT 483
Db 420 YFQSNSTLSVDRTAGSDISYDPAAGGVHTAQLAQDNTLHVHIALIDTCSVEVFGGGEA 479
Qy 484 VLSQVHFAEGDGTGISLTDGGAHFTGIVVREI 517
Db 480 VISDLIFFSNSDGLSLEVSGETAMLRSVNVSSV 513

RESULT 10
C97080
levanase/invertase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

RESULT 12
A49206
exo-beta-1
C;Species

```
Query Match      13.4%; Score 379; DB 2; Length 550;
Best Local Similarity 24.68; Pred. No. 2.7e-20;
Matches 139; Conservative 87; Mismatches 212; Indels 128; Gaps 27;

QY   39 LRAVYHMTPPSGWGLCDPQRVTTHGAY :-----QLYHLHSDQNNG-----PGWDHAST 86
    || |::|| ||: || |::| :|: |:| :|: |::|::|
```

DD LRPQLHLIPDQGMNDP-----NGMEIDRDKLWLVFQRNPDKNSINWAIFVLTGCRSIS 80
QY 87 TDGVAFTHRGIVM-PLRPDPFPVMWSGAVVGNTAGTGAGA-----VVVALTOPTDGVR 139

DB 8I KDLIMDIHGNALEPENDEDEGIFSGSVVDRNNISGFFNDSDPEQRIVATINNA - - - 130

QY :**| | :** **| | | :** **| :|** **:|** :**| :**

t40 KIQEILWSIDGGIFIALFDFVIVNIDGKAAIIFAETENAEWERDFAIRNDIAKGEWV 199

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Q	200	250	300	350	400	450	500	550	600	650	700	750	800	850	900	950	1000
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0													

[illegible][illegible]

07 310 ETEWBIATAMNNKVVAAABDDVDTDA SDCYNCONSTVBETDIAP--OBCCW-VTIISTDVA 365

301 FDCAIGIAWASNWOY-ANTVPT---ENWRSSMSIYBRYNTIKYVDVNPENYGI.TLIOKPV- 355

00 366 ALTNYVTATTTI-----PDRTVDGSAVLPWNGRAVE-----IELDIAWDT 405

```

db- 356 ---YDTKETRLNETLKTLETINEYEVNDLKDKSSFVATDFENTERNATGVFEFDLKF-- 409

```

QY 406 ATNVGISVGRSPDCTRHTNICKY-----GADLYVDRGP----- 438

```

Db      410  -TQTD LKMGS --- NMTTQFGLYIHSQTVKGSQETQLQVFDTLSTTWYIDRTTQHSFQRN 465

```

QY 439 SDLAGYSLAPYSRAAAPIDPGARSVH-LRILVDTQSVFVNAGHTVLSQQVHFAEG--D 495

Db 466 SPVFTERISTYVEKIDTTDQG--NVYTLGVVDRNILELYFNDGSIAMTNTFFFFREGKIP 523

QY 496 TGISLYTDGCPAHT--GIVVREIQ 519

Db 524 TSFEVCDSEKSFITIDELSVRELAR 549

DECEMBER 14

S27372
beta-fructofuranosidase (EC 3.2.1.26) precursor - yeast (Saccharomyces cerevisiae)

N; Alternate names: invertase; saccharase
C: species: *saccharomyces cerevisiae*

C;Date: 23-Apr-1993 #sequence_revision 23-Apr-1993 #text_change 22-Jun-1999
C;Accession: S27372: S25441

R;Hohmann, S.
submitted to the EMBL Data Library, November 1988

A;Reference number: S27372
A;Accession: S27372

A;Molecule type: DNA
A;Residues: 1-532 <HOH>

A; C/ROSS-References: EMBL:X07370; NID:94367; FIDN:CAA30437.1; FID:94368
R; Hohmann, S.; Gozalbo, D.

A;Title: Structural analysis of the 5' regions of yeast SUC genes revealed analogous

A;Accession: S25441

A; Residues: 1-74 <HO2>

A:Cross-references: EMBL:X07570

C:Genetics:

A:Gene: SUC1

A:Map position: 7R

C:Superfamily: beta-fructofuranosidase

C:Keywords: glycosidase; hydrolase

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-532/Product: beta-fructofuranosidase 1 #status predicted <MAT>

Query Match 13.0%; Score 369.5; DB 2; Length 532;
Best Local Similarity 25.1%; Pred. No. 1.3e-19;
Matches 140; Conservative 84; Mismatches 221; Indels 113; Gaps 25;

QY 19 LALIFGAVPPAARASAPGSLRAYVHTPPSGWLCDPQR--PVTHGAYQLYYLHSDQNN 76

Db 7 LFLLAGFAAKISASMTNETSDRLVHTPNKGMNDPGLWYDAKEGKWHLYFOYNPNDT 66

QY 77 G---PGWDHASTTDGVAFTTHG---TVMLPRDPFVWGSAGVGTANTAGFGAGAV-- 127

Db 67 VWGLPLFWGHATSDD---LTHWQDEPVAIAKPKNDGSGAYSGSMVIDYNNTSGFNDTDP 123

QY 128 ---VALATQPTDGVRYQEOYLWSTDGGFTFTAL-PDPVIVNTDGRAATTPAETENAE 182

Db 124 RQRCVAIWTYNTP---ESEEQYISYSLDGGYTFTEYQKNPVLA-----ANST 167

QY 183 WFRDPKIHWDTARGEWCWVIGRLR--YAAFYTSPNLRDWTLLRRNFDYPNHALGG--IECP 238

Db 168 QFRDPKRVFVWEPSSKWMITAARKSODYKIEIYSSDDLKSNKLESAP--ANEGLFGYQYCEP 225

QY 239 DLFEITA---DDGTRHWLAASMDAYGIGLPMTYA---YWTGTWDEQFADDLTPQWLWDW 293

Db 226 GLIEVPEQDPKSKHWVFIINP---GAPAGGSFNQYFVGSFNGHFEAFDQNSRVDF 282

QY 294 GMDWAAVTPSIDAPETKRLAIAMNNWKYAARDVPTDASDGYNGQNSIVRELRLAROP 353

Db 283 GRDYALQTFNTDPTYGSGALGIWASNWEYSA-FVP---SNPWRSSMLVRFPSLNT- 337

QY 354 GGYTLLSTPVAALTNVYATTTLPDRTVDGSAVLPMN-----GRAYEIELDIAM 403

Db 338 ---YQANPETELINLKA-----EPILNISAGPWSRFATNTTLTKANSYNVDLSNST 386

QY 404 DT-----ATNVGISVGRS-----PDGTRHTNLGKGADLYVDRGSPDLA 442

Db 387 GTLEFLVYAVNTQTISKVSFADLSLWFKGLEDPPEYLRMGFEVSASFFLDKRGSKYK 446

QY 443 GYSLAPY--SRAAAPIDPGARSVHLRI-----LVDTSQSVFVFNAGHTVLSQQVHFAEGD 495

Db 447 FVKENPYFTNRMSVNNQPFKSENDLSYKYVYGLLDQNLILELYFNDGDVYVSTNTYFMTTGN 506

QY 496 ---TGIS--LYTD 503

Db 507 ALGSVNMTTGVNDLIFYD 524

RESULT 15

S27373

beta-fructofuranosidase (EC 3.2.1.26) 4 precursor - yeast (Saccharomyces cerevisiae)

N:Alternate names: invertase; saccharase

C:Species: Saccharomyces cerevisiae

C:Date: 23-Apr-1993 #sequence_revision 23-Apr-1993 #text_change 22-Jun-1999

C:Accession: S27373; S25442

R:Holmann, S.

submitted to the EMBL Data Library, November 1988

A:Reference number: S27372

A:Accession: S27373

A:Molecule type: DNA

A:Residues: 1-532 <HOH>

A:Cross-References: EMBL:X07572; NID:94572; PIDN:CAA30459.1; PID:94573

R:Holmann, S.; Gozalbo, D.

Mol. Gen. Genet. 211, 446-454, 1988

A:Title: Structural analysis of the 5' regions of yeast SUC genes revealed analogous pal

A:Reference number: S25439; MUID:88216256; PMID:2835632

A:Accession: S25442

A:Molecule type: DNA

A:Residues: 1-74 <HOH2>

A:Cross-references: EMBL:X07572

A>Note: the authors translated the codon ATT for residue 7 as Asn

C:Genetics:

A:Gene: SUC4

C:Superfamily: beta-fructofuranosidase

C:Keywords: glycosidase; hydrolase

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-532/Product: beta-fructofuranosidase 4 #status predicted <MAT>

Query Match 12.9%; Score 366.5; DB 2; Length 532;
Best Local Similarity 26.2%; Pred. No. 2.2e-19;
Matches 143; Conservative 80; Mismatches 231; Indels 91; Gaps 24;

QY 21 LIFGGAAPPAARASAPGSLRAYVHTPPSGWLCDPQR--PVTHGAYQLYYLHSDQNN- 77

Db 9 LIAGFAAKISALMTNETSDRLVHTPNKGMNDPGLWYDAKEGKWHLYFOYNPNDTVM 68

QY 78 --PGWDHASTTDGVAFTTHG---TVMLPRDPFVWGSAGVGTANTAGFGAGAV--- 127

Db 69 GLPLFWGHATSND---LTHWQDEPVAIAKPKNDGSGAYSGSMVIDHNTSEFFNDTVDPRQ 125

QY 128 ---VALATQPTDGVRYQEOYLWSTDGGFTFTAL-PDPVIVNTDGRAATTPAETENAE 184

Db 126 RCVAIWTYNTP---ESEEQYISYSLDGGYTFTEYQKNPVLA-----ANSTQF 169

QY 185 RPKIHWDTARGEWCWVIGRLR--YAAFYTSPNLRDWTLLRRNFDYPNHALGG--IECPDL 240

Db 170 RPKFVWEPSSKWMITAARKSODYKIEIYSSDDLKSNKLESAP--ANEGLFGYQYCEPGL 227

QY 241 FEITA---DDGTRHWLAASMDAYGIGLPMTYA---YWTGTWDEQFADDLTPQWLWDW 295

Db 228 IEVPTQDPKSKHWVFIINP---GAPAGGSFNQYFVGSFNGHFEAFDQNSRVDF 384

QY 296 DMYAAVTPSIDAPETKRLAIAMNNWKYAARDVPTDASDGYNGQNSIVRELRL----- 350

Db 285 DYALQTFNTDPTYGSGALGIWASNWEYSA-FVPTNP---WRSSMSLVKRFSLNTEYOA 340

QY 351 RQPGGWYTLSTPVAALTN-----YVTATTLTPDRTVDGSAVLPMNGRAYEIELDIADWT 405

Db 341 NPETELINLKAEPILNISAGPWLHPASNTL-TKANSESVDLNSNSTGTLEFLVYAVNT 399

QY 406 ATNVGISV- - - - -RSPDGTHTNIGKYGADLYVDRGSPDLAGYSLAPY--SRAA 453

Db 400 TQSVSKSVFSDLSLWFKGLEDPPEYLRMGFEASASFFLDKRGNSKYKFKENPYFTNRMS 459

QY 454 APIDPGARSVHLRI-----LVDTSQSVFVFNAGHTVLSQQVHFAEGD-----TGIS- 499

Db 460 VNNQPFKSENDLSYKYVYGLLDQNLILELYFNDGDVYVSTNTYFMTTGNALGSMVTTCVDN 519

QY 500 -LYTD 503

Db 520 LFYID 524

Search completed: May 6, 2003, 12:35:49

Job time : 31.9442 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:23:52 ; Search time 10.4798 Seconds
(without alignments)
2054.066 Million cell updates/sec

Title: US-09-868-328B-4

Perfect score: 2837

Sequence: 1 MTPAIRRAVLQAGAGALA.....LYTDGGPAHFTGIVVREIGQ 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
1	556	19.6	677	1	SACC_BACSU	P05656 bacillus su
2	382.5	13.5	1423	1	FRUA_STRMU	Q03174 streptococc
3	379	13.4	550	1	INV1_HANAN	P40912 hansenula a
4	369.5	13.0	532	1	INV1_YEAST	P10594 saccharomyc
5	366.5	12.9	532	1	INV4_YEAST	P10596 saccharomyc
6	365.5	12.9	555	1	INV1_KLUMA	P28999 kluyveromyc
7	359.5	12.7	532	1	INV2_YEAST	P00724 saccharomyc
8	354	12.5	477	1	CSCA_ECOLI	P00714 escherichia
9	348	12.3	581	1	INV1_SCHPO	O59852 schizosacch
10	330	11.6	476	1	RAFD_ECOLI	P16553 escherichia
11	320	11.3	533	1	INV_DEBOC	P24133 debaryomyce
12	312	11.0	590	1	INVA_MAIZE	P49174 zea mays (m
13	276	9.7	512	1	INVA_ZYMO	P35636 zymomonas m
14	274.5	9.7	636	1	INVA_LYCES	P29000 lycopersico
15	260.5	9.2	640	1	INV1_CAPAN	P33761 capsicum an
16	249	8.8	642	1	INVA_VICFA	Q43857 vicia faba
17	248	8.7	651	1	INVA_PHAVU	O24509 phaseolus v
18	246.5	8.7	465	1	SCRB_KLEPN	P27217 klebsiella
19	246.5	8.7	555	1	INV1_PEA	Q43089 pisum sativ
20	246	8.7	661	1	INV1_DAUCA	P80065 daucus caro
21	243	8.6	670	1	INV1_MAIZE	P49175 zea mays (m
22	242	8.5	592	1	INV1_DAUCA	P26792 daucus caro
23	242	8.5	649	1	INVA_PHAAU	P29001 phaseolus a
24	239.5	8.4	511	1	SCRB_ZYMO	P22632 zymomonas m
25	238.5	8.4	432	1	BFRA_THEMA	Q33833 thermotoga
26	237.5	8.4	494	1	SCRB_STAXY	Q05936 staphylococ
27	232.5	8.2	484	1	SCRB_VIBAL	P13394 vibrio algi
28	222	7.8	454	1	SCRB_STRMU	P13522 streptococc
29	214	7.5	466	1	SCRB_SALTY	P37075 salmonella
30	213	7.5	473	1	SCRB_LACLA	Q04937 lactococcus
31	206	7.3	501	1	SCRB_PEDPE	P43471 pediococcus
32	204.5	7.2	583	1	INV3_DAUCA	Q39693 daucus caro
33	197.5	7.0	480	1	SCRB_BACSU	P07819 bacillus su

34 187.5 6.6 592 1 INV2_DAUCA
35 121 4.3 1120 1 STFR_ECOLI
36 117 4.1 451 1 Y483_MYCTU
37 117 4.1 3491 1 ERY1_SACER
38 116.5 4.1 984 1 EPA3_RAT
39 114.5 4.0 890 1 NIA2_PHAVU
40 114 4.0 1132 1 VHSJ_LAMBD
41 113.5 4.0 909 1 NIA_PETHY
42 113 4.0 296 1 ALYS_BPDP1
43 112 3.9 916 1 NIAL_ORYSA
44 111.5 3.9 752 1 HISA_PROAC
45 110.5 3.9 983 1 EPA3_HUMAN

Q39692 daucus caro
P76072 escherichia
Q11149 mycobacteri
Q03131 saccharopol
O08680 rattus norv
P39866 phaseolus v
P03745 bacterioph
P36859 petunia hyb
O03979 bacterioph
P16081 oryza sativ
Q59634 propionibac
P23320 homo sapien

ALIGNMENTS

RESULT 1

SACC_BACSU
ID SACC_BACSU STANDARD; PRT; 677 AA.
AC P05656;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Levanase precursor (EC 3.2.1.65) (2,6-beta-D-fructan
DE fructanohydrolase).
GN SACC.
OS Bacillus subtilis.
OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87286401; PubMed=3112519;
RA Martin I., Debarbouille M., Ferrari E., Klier A., Rapoport G.;
RT "Characterization of the levanase gene of Bacillus subtilis which
RT shows homology to yeast invertase.";
RL Mol. Gen. Genet. 208:177-184(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RC MEDLINE=88067786; PubMed=3120151;
RA Schoergendorfer K., Schwab H., Lafferty R.M.;
RT "Nucleotide sequence of a cloned 2.5 kb PstI-EcoRI Bacillus subtilis
RT DNA fragment coding for levanase.";
RL Nucleic Acids Res. 15:9606-9606(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RC MEDLINE=97286548; PubMed=9141695;
RA Parro V., San Roman M., Galindo I., Purnelle B., Bolotin A.,
RA Sorokin A., Mellado R.P.;
RT "A 2391 bp region of the Bacillus subtilis genome comprising genes
RT located upstream and downstream of the lev operon.";
RL Microbiology 143:1321-1326(1997).
RN [4]
RP SEQUENCE OF 1-10 FROM N.A.
RX MEDLINE=90355183; PubMed=2117666;
RA Martin-Verstraete I., Debarbouille M., Klier A., Rapoport G.;
RT "Levanase operon of Bacillus subtilis includes a fructose-specific
RT phosphotransferase system regulating the expression of the operon.";
RL J. Mol. Biol. 214:657-671(1990).
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 2,6-beta-D-
CC fructofuranosidic linkages in 2,6-beta-D-fructans (Levans)
CC containing more than 3 fructose units.
CC -1- INDUCTION: IN CONTRAST WITH LEVANSUCRASE AND SUCRASE, LEVANASE
CC SYNTHESIS IS NOT INDUCIBLE BY SUCROSE; NO INDUCER IS KNOWN FOR
CC LEVANASE.
CC -1- MISCELLANEOUS: LEVANASE CANNOT BE DETECTED IN THE WILD-TYPE
CC B. SUBTILIS BUT IS MOSTLY SECRETED INTO THE CULTURE MEDIUM BY SACL
CC MUTANTS, ESPECIALLY AT THE END OF THE EXPONENTIAL GROWTH PHASE.
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X05649; CAA29137.1; ALT_INIT.
CC EMBL; Y00485; CAA68542.1; -.
CC EMBL; X92868; CAA63465.1; -.
CC EMBL; X56098; CAA39581.1; -.
CC EMBL; Z99117; CAB14645.1; -.
CC PIR; S06353; S06353.
CC PIR; A27286; A27286.
CC Subtilist; BG10320; sacC.
CC InterPro; IPR001362; GH_32.
CC Pfam; PF00251; Glyco_hydro_32; 1.
CC PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Signal; Complete proteome.
FT CHAIN 1 24 POTENTIAL.
FT ACT_SITE 25 677 LEVANASE.
FT CONFLICT 49 49 BY SIMILARITY.
FT CONFLICT 658 658 O -> L (IN REF. 2).
SQ SEQUENCE 677 AA; 75951 MW; 80FD6B0A5EE7F525 CRC64;
Query Match 19.6%; Score 556; DB 1; Length 677;
Best Local Similarity 32.2%; Pred. No. 9.3e-33;
Matches 160; Conservative 74; Mismatches 217; Indels 46; Gaps 18;
QY 40 RAVYHMTPPSGWMLCDPQRPVTHHGYQLYYLHSDQNN--GPGGWDHASTTDCGVAFTTHGCT 97
DB 35 RQYHFTPPANNMNDGVYAGETHLYFYHPYGLQWPMHWGHAVSKDLVTWPH--L 92
QY 98 VMLPRD--FPVWSGSAVGTANTAGFAG---AVVALATQPTDGVKRYQEQYLYWSTDG 152
DB 93 PVALYPDEKGTIFSGSAVDKNTSGFTGKEKPLVAITYTDREG---HQVSIAYSNDK 149
QY 153 GTTFTALP--DPVIVNTDGRAATTAPAEIENAEWFRDPKIHWDTARGEWVCVIGRLRYAAPY 211
DB 150 GRTWTKYAGNPVNP-----PGKGD---FRDPKVFWEKEKKWVLAAGDRILIY 197
QY 212 TSPNLRDWTLLRNEDYPNHALGI--ECPDLFEITAD--DGRHWVLAASMDAYGICLPM 268
DB 198 TSKNLKQWITYASEFGQDQSGHGWECPLDFELPDVGNPNQKKNVQSVGNCAVSGSG 257
QY 269 YAYWTGTWDEGFHADDTLPQ--WLDWGDWTAAYTWPISIDAPETKRLAIANNWKKYAA 326
DB 258 MQYFVCGDFDGTHEKNENPNKVLWTDYGRDFAAYSWSDIPSTDRLRLGWNWNNQY-A 316
QY 327 RDVPTDASDGYNGQNSIVRELRLAROPGGWYTLTSTPVAALTYNYVATTTLPDRTVDGSA 386
DB 317 NDVPTSP---WRSATSIPELKLKAFTEG--VRVQVTPVKLELETIRGTSKKKNLTISPAS 372
QY 387 ---VLPWNCRAVEI--ELDIANDTATNVCISVGRSPDGTTRHNTIGKYADLYVDGRPSDL 441
DB 373 HNVLAGQSGDAYEINAEFVPGSAAEFGFKVYRTGENQTKYGYDRRNKALFEVDRSEGN 432
QY 442 AGYSILA-PYSRAAPIDPCARSVHLRLVDTQSVFVFNAGHTVLVSQOVHFAEGDGTISL 500
DB 433 DTFNFAFTNGKETAPLKPVNGKVKLRIFFVDRSSVEVFGNDGKQVITDILPDRSSKGLEL 492
QY 501 YTDGGPAHTGIVREI 517
DB 493 YAANG-----GVKVKSL 504
RESULT 2
FRUA_STRMU
ID FRUA_STRMU STANDARD; PRT; 1423 AA.
AC Q03174;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
```

```
DE Fructan beta-fructosidase precursor (EC 3.2.1.80) (Exo-beta-D-
DE fructosidase) (Fructanase).
GN FRUA.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=93014178; PubMed=1398976;
RA Burne R.A., Penders J.E.C.;
RT "Characterization of the Streptococcus mutans GS-5 frua gene encoding
RL exo-beta-D-fructosidase.";
RL Infect. Immun. 60:4621-4632(1992).
CC -!- FUNCTION: THIS PROTEIN IS A FRUCTANASE ENZYME WHICH DEGRADES
CC LEVANS AND INULINS TO FRUCTOSE AND ALSO CLEAVES SUCROSE INTO
CC GLUCOSE AND FRUCTOSE AND CAN THEREFORE FUNCTION AS AN
CC EXTRACELLULAR INVERTASE.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal-non-reducing 2,1- and
CC 2,6-linked beta-D-fructofuranose residues in fructans.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan
CC by an amide bond (Potential).
CC -!- INDUCTION: BY SUCROSE, FRUCTAN SUBSTRATES AND FRUCTOSE.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U78296; AAA26889.1; -.
CC InterPro; IPR003343; Big_2.
CC InterPro; IPR001362; GH_32.
CC InterPro; IPR001899; Gram_pos_anchor.
CC Pfam; PF00251; Glyco_hydro_32; 1.
CC Pfam; PF02368; Big_2; 1.
CC TIGRfams; TIGR01167; LPXTG_anchor; 1.
CC PROSITE; PS00609; GLYCOSYL_HYDROL_F32; FALSE_NEG.
CC PROSITE; PS50847; GRAM_POS_ANCHORING; FALSE_NEG.
KW Hydrolase; Glycosidase; Cell wall; Peptidoglycan-anchor; Signal.
FT SIGNAL 1 39 POTENTIAL.
FT CHAIN 40 1391 FRUCTAN BETA-FRUCTOSIDASE.
FT PROPEP 1392 1423 REMOVED BY SORTASE (POTENTIAL).
FT ACT_SITE 458 458 BY SIMILARITY.
FT DOMAIN 867 871 INVOLVED IN BINDING OF SUGARS WITH
FT BETA-(2,6) LINKAGES OR BINDING OF
FT MOLECULAR WEIGHT FRUCTANS (BY
FT SIMILARITY).
FT SITE 1388 1392 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1391 1391 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 1423 AA; 158668 MW; 8E574715F4E72A8A CRC64;
Query Match 13.5%; Score 382.5; DB 1; Length 1423;
Best Local Similarity 28.2%; Pred. No. 7.6e-20;
Matches 148; Conservative 61; Mismatches 193; Indels 123; Gaps 27;
QY 40 RAVYHMTPPSGWMLCDPQRPVTHHGYQLYY-LHSDQNNPGGWDHASTTDCGVAFTTHGTV 98
DB 444 RDQYHYSVKDGNWDPNGLVYNGVYHLPFHQFYDDTKGPMHWAHATSTDLI-----HWKE 499
QY 99 MPLR--PDFP--VWMSGVAVGTANTAGF---GAGAVVALATQPTDGVKRYQEQYLYWSTD 151
DB 500 EPIAFYPSDNGVWMSGVAVGTANTAGF---GAGAVVALATQPTDGVKRYQEQYLYWSTD 555
QY 152 GGFTFTALPDVIVNTDGRAATTAPAEIENAEW-----FRDPKI--HWDTARGEWVCV 201
DB 556 ECKTWQKY-DRIV-----ADWSNDPLQNDQFDPKVFHWN---NQWFWV 595
QY 202 I--GLRLYAAFTYSPNLRDWTLLRNEDYPN--HALGIECPDLFEITADDTGRHWVLAAS 258
```



```

Db 596 LAGGFLR---YSSNNLKDWKVEST--YPLDHT-----ECPDMYPIVANDGVKLKVLRSRG 646
Qy 259 DAYGIGLPMYAYWTGWDGEFHADDL--TPQWLDGWDYAAVW-----PS 305
Db 647 REYKVG--DFKQVDGKW--TFIADDAKDKQDQVNFEGKDSYAAWTVYVDFGTETRPT 700
Qy 306 IDAPETKRLAIAMNNWYKAAARDVPTDASDGYNGQNSIVRELRLARQPGGWYTLTSTPVA 365
Db 701 I-----PKLTEVNMWNTWEDYCNLVAADVTDGQDFNGFTNLNLDGLINE-NGOYILTQTPVK 755
Qy 366 ALTNVVTATTL--PDRVDSGAVL--PWNGRAYEIELDIAMDATNATNGVSGVRSPDGTR 421
Db 756 AYDSLRDYNTALHFQKDVTVTDANNTLLKDFKGSYIVSHFRPDEKTT-----802
Qy 422 HTNICKYGADLVVDRGSDLAGYSLAP-----YSRAAPIDP-----GARS 462
Db 803 -----KVGFNLRVGMQKATKYIYDLOTLSIDRSQSGTILSAFAKVNQHVTRNADGS 857
Qy 463 VHLRLVDTQSVYEVFNAGHTVLSQQVHFAEGDTGISLYTDGGA 507
Db 858 IDLHIYVDRAVSEVFESKNTVAGANQIFPNPEAVGASIIIEGKA 902

RESULT 3
INV1_HANAN
ID INV1_HANAN STANDARD; PRT; 550 AA.
AC P40912;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Invertase precursor (EC 3.2.1.26) (Beta-fructofuranosidase 2)
DE (Saccharase).
GN INV1.
OS Hansenula anomala (Yeast) (Candida pelliculosa).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4927;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96163052; PubMed=8595669;
RA Perez J.A.; Rodriguez J.; Rodriguez L.; Ruiz T.;
RT "Cloning and sequence analysis of the Invertase gene INV 1 from the
RT yeast Pichia anomala."
RL Curr. Genet. 29:234-240(1996).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X80640; CAA56684.1; -
CC PIR; S48812; S48812.
CC InterPro; IPR001362; GH_32.
CC Pfam; PF00251; Glyco_hydro_32; 1.
CC PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
CC Hydrolyase; Glycosidase; Glycoprotein; Signal.
CC SIGNAL 1 22 POTENTIAL.
CC CHAIN 23 550 INVERTASE.
CC FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 237 237 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
...FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 63188 MW; 192BCD722455E121 CRC64;

Query Match 13.4%; Score 379; DB 1; Length 550;
Best Local Similarity 24.6%; Pred. No. 4.3e-20;
Matches 139; Conservative 87; Mismatches 212; Indels 128; Gaps 27;

Qy 39 LRAVYHMPPPSGWLCDPQRPVTHGAY-----QLYVLHSDQNNG-----PGWHDHAST 86
Db 27 LRQJHLHPDQGMNDP-----NMFYDRDKLHHVVFQHNPKKSIWATPVTHGHS 80
Qy 87 TDGVAFTHHGTVM-PLRPDPFVWSGSAVVGTAANTAGFCAGA-----VVALATQPTDGYR 139
Db 81 KDLLTWYHGNALPENDEDEGIFSGSVVVDNRNNTSGFFNDSTDPEQRIVAIYTNA----136
Qy 140 KYQEOYLWSTDGGTFTALPDPPVIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGEW 199
Db 137 QLQTOEIAIYSLDKYSFKYDQNPVINV-----NSSQQRDPKVLWHDESNOWI 184
Qy 200 CVIGRLR--YAAFYTSPLNRDMWTLRRNFDYPNHALGGIECPDLFEITA-----DDGTRHW 252
Db 185 MVVAKTQEFKVOIYGPDLKKWDLKSNFTSNGYLGFEQECPLFLPIENPLNDIVTSKW 244
Qy 253 VLAASMDAYGIGLPM---TYAYWTGTWDEQPHADDLTQWLDGWDWYAAVTFPSIDAP 309
Db 245 VLLALNP---GSLGGSGINEYFIDGDTTFHPDGGATRFMDICKDFYAFQSFQNTF-P 300
Qy 310 ETKRLAIAMNNWYKAAARDVPTDASDGYNGQNSIVRELRLAR---OPGGW-YTLTSTPVA 365
Db 301 EDGALGLAWASNMWQY-ANTVPT---ENWRSSMSLVNRYTLKYVDVNPENYGLTLTKQPV- 355
Qy 366 ALTNVVTATTL-----PDRVDSGAVLPWNGRAYE-----LELDIAMD 405
Db 356 ----YDTETRLNETLKTLETINEYEVNDLKLKSSFVATDENTERNATGVEFDLKF---409
Qy 406 ATNNGISVGRSPDGRTHNIGKY-----GADLYVDRGP-----438
Db 410 -TQTDLKMGS---NMTTQFGLYIHSQTVKGSQETLQLVFDLTSTTWYIDRTTQHSFORN 465
Qy 439 SDLAGYSLAPYSRAAPIDPGARSVH-LRLVDTQSVYEVFNAGHTVLSQQVHFAEG--D 495
Db 466 SPVFTERTISTYVEKIDTTDQG--NVYTLGVVDNRNILELYFNDGSIAMTNTFFREGKIP 523
Qy 496 TGISLYTDGGAHFT--GIVVREIGO 519
Db 524 TSFEVVCDSKSFITIDELSVRELAR 549

RESULT 4
INV1_YEAST
ID INV1_YEAST STANDARD; PRT; 532 AA.
AC P10594;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Invertase 1 precursor (EC 3.2.1.26) (Beta-fructofuranosidase 1)
DE (Saccharase).
GN SUC1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-73 FROM N.A.
RX MEDLINE=88216256; PubMed=2835632;
RA Hohmann S.; Gozalbo D.;
RT "Structural analysis of the 5' regions of yeast SUC genes revealed
RT analogous palindromes in SUC, MAL and GAL."
RL Mol. Gen. Genet. 211:446-454(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA Hohmann S.;

```


[illegible]

homologous inulinase promoter in Kluyveromyces marxianus.",
Appl. Microbiol. Biotechnol. 40:309-317(1993).
[3]
SEQUENCE OF 24-43.
MEDLINE=922304047; PubMed=2135869;
Rouwenhorst R.J., Hensing M., Verbakel J., Scheffers W.A.,
van Dijken J.P.;
"Structure and properties of the extracellular inulinase of
Kluyveromyces marxianus CBS 6556.";
RT Appl. Environ. Microbiol. 56:3337-3345(1990).
CC -!- FUNCTION: HAS BOTH INULASE AND INVERTASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 2,1-beta-D-fructosidic
linkages in inulin.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X57202; CAA40488.1; -;
DR EMBL: X68479; CAA48500.1; -;
DR PIR: S17502; S17502.
DR InterPro: IPR001362; GH_32.
DR Pfam: PF00251; Glyco_hydro_32; 1.
DR PROSITE: PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 15 OR 16, OR 17 (POTENTIAL).
FT PROPEP 16 23 POTENTIAL.
FT CHAIN 24 555 INULINASE.
FT ACT_SITE 53 53 BY SIMILARITY.
FT CARBOHYD 32 32 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 122 122 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 123 123 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 129 129 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 220 220 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 226 226 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 255 255 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 385 385 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 393 393 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 490 490 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 549 549 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 3 3 F -> L (IN STRAIN CBS 6556).
FT VARIANT 165 165 E -> Q (IN STRAIN CBS 6556).
FT VARIANT 168 168 T -> S (IN STRAIN CBS 6556).
FT VARIANT 251 251 V -> A (IN STRAIN CBS 6556).
FT VARIANT 252 252 D -> DS (IN STRAIN CBS 6556).
FT VARIANT 299 299 G -> D (IN STRAIN CBS 6556).
FT VARIANT 479 479 T -> N (IN STRAIN CBS 6556).
FT CONFLICT 24 24 D -> S (IN REF. 3).
FT CONFLICT 43 43 H -> Y (IN REF. 3).
SQ SEQUENCE 555 AA; 62213 MW; 723BAADC3BF0907 CRC64;

Query Match 12.9%; Score 365.5; DB 1; Length 555;
Best Local Similarity 27.0%; Pred. No. 4.1e-19;
Matches 132; Conservative 72; Mismatches 181; Indels 103; Gaps 24;

QY 40 RAVYHMTTPSGWLCDDPQRPVTHGA-----YOLYVYHSDQN---NGPGWGHASTD 88
DB 39 RPSVHFTPSHGWNNDP-----NGLWYDAKEDWHLYIQYNPATNGTPTLYWGHAVSKD 92

QY 89 GVAFTHHGTVM-PLRPDPFPWWSGSAVVGTAFTAGF-----GAGAVVALATQPTDGV 138
DB 93 LFSWTYDYGASLGSGDDACAFSGSMVIDYNTSGFNSVDPQRAVAVWTLKSGPS--- 149
QY 139 RKYQYLYXWSTDGGFTFTALPDVIVNTDGRAATTPAEINAEWFRDPKIHWDAR--- 195
"Expression of an alpha-galactosidase gene under control of the

Db 150 ---QAQHSISLGGYTFEHTYDNAVLDI-----NSSNFRDPRKVFVHEGENGE 194
QY 196 -GMVVCVIGRLY--AAFYTSPLNRDLTTLRRNFDPYHNLGEGIECPDLFEITAD---DGT 249
Db 195 DGRWIMAVAESQVFSVLFSYSSPNLKNWLTSTHGHGTGTQYECPLGVKVPYDSSVVDSS 254
QY 250 RH-----WVLAASMDVAG-TGLPWTYAYTGTWGDGEQFHADDLTLPQWLWDGNDWYAA 300
Db 255 NSSDSKPDASAWLFSINPGGPGGSVT-QYFVGDFNGTHFTPDGTQTRFLDMGKDYAL 313
QY 301 VTPWPSIDAPETKRL-ATAMNNKYYAARDVPTDASDGYNCGNSIVRELRL---AROPGGW 356
Db 314 QTF--FNTPEKDYVGIWASNNQY-AQAQPTOP---WRSSMLVROFTLUKDFSTPNNSA 367
QY 357 YTLT-STPVAALNTYVATTTLTDPRTVDGSAVLPPWNGRAVEIELDTAWDTATNNGISVGR 415
Db 368 DVLNSQPV---LNYDALRNGTYSITNTVTSSENGK--KIKLD----- 407
QY 416 SPQTRHTNGKYGADLYVDGRGSDLAGYSLAPYSRAAAPIDPGARSHLRILVDQTSVE 475
Db 408 NPSSGSEFHE-----YFNGSPDIKSNVFADSLFLYFKGNDD--NEYLRLGYETNGGA 459
QY 476 VEYNAGHT 483
Db 460 FFLDRGHT 467
RESULT 7
INV2_YEAST STANDARD; PRT; 532 AA.
AC P00724;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Invertase 2 precursor (EC 3.2.1.26) (Beta-fructofuranosidase 2)
DE (Saccharase)
GN SUC2 OR YII162W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83168934; PubMed=6300785;
RA Taussig R., Carlson M.;
RT "Nucleotide sequence of the yeast SUC2 gene for invertase.";
RL Nucleic Acids Res. 11:1943-1954(1983);
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moulé T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-30 FROM N.A.
RX MEDLINE=85137467; PubMed=6396505;
RA Sarokin L., Carlson M.;
RT "Upstream region required for regulated expression of the glucose-
RT repressible SUC2 gene of Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 4:2750-2757(1984).
RN [4]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=87064535; PubMed=3537718;
RA Kaiser C.A., Botstein D.;
RT "Secretion-defective mutations in the signal sequence for
RL Saccharomyces cerevisiae invertase.";
RL Mol. Cell. Biol. 6:2382-2391(1986).
RN [5]
RP SEQUENCE OF 21-532, AND CARBOHYDRATE-LINKAGE SITES.

RX MEDLINE=88213364; PubMed=3284881;
RA Reddy V.A., Johnson R.S., Biemann K., Williams R.S., Ziegler F.D.,
RA Trimble R.B., Malsey F.;
RT "Characterization of the glycosylation sites in yeast external
RT invertase. I. N-linked oligosaccharide content of the individual
RT sequons.";
RL J. Biol. Chem. 263:6978-6985(1988).
RN [6]
RP ACTIVE SITE ASP-42.
RX MEDLINE=90293006; PubMed=2113524;
RA Reddy V.A., Malsey F.;
RT Identification of an active-site residue in yeast invertase by
RT affinity labeling and site-directed mutagenesis.";
RL J. Biol. Chem. 265:10817-10820(1990).
CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -I- PTM: THE ENZYME EXISTS IN TWO STATES: A GLYCOSYLATED EXTERNAL
CC FORM, OR A NON GLYCOSYLATED INTERNAL FORM.
CC -I- MISCELLANEOUS: THE SEQUENCE OF THE SECRETED FORM OF THE ENZYME IS
CC SHOWN.
CC -I- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 246921; CAA87030.1;
DR EMBL; V01311; CAA24618.1;
DR EMBL; K03294; AAA35127.1;
DR EMBL; M13627; AAA35129.1;
DR PIR; A00899; IFBY
DR PIR; A27748; A27748.
DR GlycosuiteDB; P00724;
DR COMPLEVEAST-2DPAGE; P00724;
DR SGD; S0001424; SUC2.
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Multigene family; Signal.
FT SIGNAL 1 19
FT CHAIN 20 532
FT ACT_SITE 21 532
FT ACT_SITE 42 42
FT CARBOHYD 23 23
FT CARBOHYD 64 64
FT CARBOHYD 97 97
FT CARBOHYD 111 111
FT CARBOHYD 118 118
FT CARBOHYD 165 165
FT CARBOHYD 266 266
FT CARBOHYD 275 275
FT CARBOHYD 356 356
FT CARBOHYD 369 369
FT CARBOHYD 384 384
FT CARBOHYD 398 398
FT CARBOHYD 512 512
FT MUTAGEN 42 42
FT CONFLICT 409 409
SQ SEQUENCE 532 AA; 60639 MW; 7D8AB33E6772B775 CRC64;
Query Match 12.7%; Score 359.5; DB 1; Length 532;
Best Local Similarity 26.0%; Pred No. 1,le-18;
Matches 143; Conservative 81; Mismatches 228; Indels 99; Gaps 25;
QY 19 LALIFGAVPPAARASAPGSLRAVYHMTPPSGWLCDDP-----QRPVTHGAVQLYLLHS 72
Db 7 LFLAGFAAKISASMTNETSDRPLVHFTPNKGMNDPGLWYDEKDAKWH---LYFQYN 62
QY 73 DQN---NGPGGWDHASTTGGVAFTHHG-TVMPLRPFVWGSAAVVGTAAGFGAGAV- 127

```
Db 63 PNDVTGMLFPGHATSDDLTNWEDQIAIAPKRNDSDGAFSGSMVVDYNTSGFFNDRTD 122
Qy 128 -----VALATOPTDGRVYQEOYLWSTDDGGFTTAL-PDPVIVNTDGRAATTPAEIEA 181
Db 123 PRQRCVATWYNTP---ESEEQYISYLDGGYTFTEYQKNPVA-----ANS 166
Qy 182 EWFDPKTHMDTARGWCVIGRLR--YAAFTSPNLRDWTLLRRNFDYPNHALGG--TEC 237
Db 167 TOFRDPKVFWEPSQKWITAAKSODYKIEIYSSDDLKSWKLESAF--ANEGLGQYEC 224
Qy 238 PDLFEITA--DDGTRHWLAASMDAYGIGLPMTYA--YWTGTWDEGDFHADDLTPOWLD 292
Db 225 POLIEVPEODPSKYWMFISINP---GAPAGGSFNOYFVGSFNGTFHEAFDQNSRVVD 281
Qy 293 WGDWYAAATWPSIDAPETKRLAIAMNNKYAARDVPTDASDGYNGQNSIVRELRL--- 349
Db 282 FGKDYALQTFNTDPTVGSALGIAWASWEISA-FVTNP---WRSSMLVRKESLNT 337
Qy 350 --ARPGGWYLLTSPVAALN-----YVTATTLPTORT--VDGSVLPWNGRAYEIEL 399
Db 338 YOANPETELINLKARPILNISNAGPWSRFATNTTLTKANSYNVD-----LSNSTGTLEFEL 393
Qy 400 DIANDTATNVSIGV-----RSPDGRHTNIGKYGADLYDRGPSDLAGYSLAPY 449
Db 394 YVAVNTTQTKSVPFADLSLMFKGLEDPPEYLRMGFEVSASFLLDRGNSKYKFKENPY 453
Qy 450 --SRAAAPIDPGARSVHLRI-----LVDTQSVFEPVNAVAGHTVLSQOVHFAEGD----- 495
Db 454 FTRNSVNNQPFKSENDLSYKYVGLLDONILELFNDGDVVSTWYFTWYTGALGSVM 513
Qy 496 -TGIS--LYTD 503
Db 514 TTGVNDLEYID 524

RESULT 8
CSCA_ECOLI
ID CSCA_ECOLI STANDARD; PRT; 477 AA.
AC P40714;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sucrose-6-phosphate hydrolase (EC 3.2.1.26) (Sucrase) (Invertase).
GN CSCA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EC3132;
RA Bockmann J.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS ENZYME ENABLES THE BACTERIA TO METABOLIZE SUCROSE
CC AS SOLE CARBON SOURCE.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -!- PATHWAY: Sucrose metabolism
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X81461; CAA57219.1; -
DR InterPro: IPR001362; GH_32.
DR Pfam: PF00251; Glyco_hydro_32; 1.
```

```
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Carbohydrate metabolism.
FT ACT_SITE 39 BY SIMILARITY.
SQ SEQUENCE 477 AA; 54363 MW; 48E6622AEA380EBD CRC64;

Query Match 12.5%; Score 354; DB 1; Length 477;
Best Local Similarity 27.2%; Pred. No. 2.3e-18;
Matches 135; Conservative 67; Mismatches 206; Indels 88; Gaps 23;

Qy 43 YHMTPPSGWGLDQDPQPVTHTGAYQLYLIHSDQNN--GPGMDHASTTGGVAFTHHGTVP 100
Db 28 FHLAPPAGWMDPNGLIWFNDRIYHAFYOHHPMSEHWGPMHGHATSDDMIHWQHE--PIA 85
Qy 101 LRPDEP-----VWSSGSAVYVGTANTAGGAGAVVALATOPTDGVRRYQEOYLWSTDDGGFT 155
Db 86 LAPGDENDKDCGFCSSGSAVDNGLSLIYTHVWLDGAGNDDAIR--EVOCLATSRD--- 140
Qy 156 FTALDPPVIVNTDGRAATTPAEIEAENAFPRDKIHWDTARGEWVCV-----IGRLRYAIFY 211
Db 141 -----IHFEKQGVILTPP---EGIMHFRDPKV--WREADTWMMVVGAKDPGNTGQILLY 189
Qy 212 TSPNLRDWTLLRNFDYPNHALGG---TECPDLREITADGGTRHWVLA--SMDAYGIGL 265
Db 190 RGSSUREWTFDRVL---AHADAGESYMWECDFPFL-----GDQHYLMFSPOGMNAEG--- 239
Qy 266 PMTYAYWT-----GTWDGEGDFHADDLTPOWLDWGDWYAAVTPSIDAPETKRLAIA 317
Db 240 ---YSYRNRFQSGVIVPGWMSPCRLFAQSGHGTDLNHDGFYAP---QSFVAKDGRRIYIG 293
Qy 318 WNNWKYAADPYDPTDASDGYNGQNSIVRELRLARPGGWYLLTSPVAALNYYVTATTL 377
Db 294 WMDWE---SPMPS--KREGWAGCMTLAREL---SESNG--KLLQRPVHEAESLROQHOSI 344
Qy 378 PDRTVDGSVLPWNGRAYEIELDIATW--DTATNVGISVGRSPDGTGTRHTNIGKYGADLYVD 435
Db 345 SPRTISNKYVLOENAAQAVEIQWALKNSDAEHYGLQGA-----GMRLYID 391
Qy 436 RGPDSLACYSLAPYSR-----AAAPIDPGARSVHLRIYLDVTSQSVFVFNAGHTVLSQOVH 491
Db 392 NQSERLVLRYYPHENLDGYSRISPL-POGDMALRIFDITSSVEVFINDGEAVMSSRIYP 450
Qy 492 AEGDTGISLYTDGGPA 507
Db 451 QPEERELSLYASHGVA 466

RESULT 9
INV1_SCHPO
ID INV1_SCHPO STANDARD; PRT; 581 AA.
AC O59852; P78891;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Invertase precursor (EC 3.2.1.26) (Beta-fructofuranosidase)
DE (Saccharase).
DE INV1 OR SPCC191.11.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98204792; PubMed=9535817;
RA Tanaka N., Ouchi N., Mukai Y., Ozaka Y., Ohtani Y., Tabuchi M.,
RA Bhyuiyan M., Fukui H., Harashina S., Takegawa K.;
RT "Isolation and characterization of an invertase and its repressor
RT genes from Schizosaccharomyces pombe.";
RL Biochem. Biophys. Res. Commun. 245:246-253(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
```

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Mouton S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voiclaert G., Aert R., Robben J., Grymonprez B.,
RA Weltyens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,
RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe";
RN Nature 415:871-880(2002).
RN [3]
RP SEQUENCE OF 88-581 FROM N.A.
RC STRAIN=PR745;
RX MEDLINE=90253381; PubMed=2187435;
RA Moreno S., Sanchez Y., Rodriguez L.;
RT "Purification and characterization of the invertase from
RT Schizosaccharomyces pombe";
RL Biochem. J. 267:697-702(1990).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -!- PTM: THE CARBOHYDRATE CONTENT OF THIS PROTEIN IS 67%. THIS IS
CC COMPOSED OF EQUIPOLAR AMOUNTS OF MANNOSE AND GALACTOSE. THERE IS
CC ALSO A SMALL AMOUNT OF GLUCOSAMINE PRESENT.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011433; BAA25684.1;
DR EMBL; AL049644; CAB41057.1;
DR EMBL; D89242; BAA13903.1;
DR InterPro: IPR001362; GH_32.
DR Pfam: PF00251; Glyco_hydro_32; 1.
DR PROSITE: PS00609; GLYCOSYL_HYDROL_F32; 1.
DR KW Hydrolase; Glycosidase; Glycoprotein; signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 581 INVERTASE.
FT ACT_SITE 97 97 BY SIMILARITY.
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (PROBABLE).

FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 446 446 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 569 569 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CONFLICT 186 189 MISSING (IN REF. 3).
FT CONFLICT 190 190 H -> Y (IN REF. 3).
FT CONFLICT 195 195 A -> P (IN REF. 3).
FT CONFLICT 198 198 Q -> L (IN REF. 3).
FT CONFLICT 205 205 I -> N (IN REF. 3).
FT CONFLICT 243 243 I -> M (IN REF. 3).
FT CONFLICT 246 246 M -> L (IN REF. 3).
FT CONFLICT 256 256 S -> P (IN REF. 3).
SQ SEQUENCE 581 AA; 64407 MW; 70206A6CD1F27EC4 CRC64;
Query Match 12.3%; Score 348; DB 1; Length 581;
Best Local Similarity 24.9%; Pred. No. 7.9e-18;
Matches 132; Conservative 82; Mismatches 226; Indels 90; Gaps 22;
QY 28 PPAARASAPGSL-----RAVYHMTFPGSGWLCDDPQRPVTHGAYQLYY 69
DB 53 PPVNTTAPNGTCLGNYNEPLPSGYNATDRPKIHFTSPSGFMNDPNGLVYTGGVYHMF 112
QY 70 LHSQD--NNGPGGDHASTTDDGVAF-----THGTVMPLRDPDFPWSGVAVGTA 117
DB 113 QVSPKTLTAGVHWGHTVSKDLIHENYPIAYDPDEHENGVLSL-----PFGSAVVDVH 167
QY 118 NTAFGAG-----AVVALATQPTDGVKRYQBYLWSTDGGFTFTALPDPVIVNTDGRA 171
DB 168 NSSGLFSNDTPEERIVLYTDHWTGVA--ERQAIAYTTDGGYTFK-----KY 213
QY 172 ATPPAEIAENAEWFRDPKIHWDTARGECWCVIGELRY--AAFTSPNLRDWTLRNFDYPN 229
DB 214 SGNPVLIDINSLOFRDPKIVDFDANRWVMIVAMSONYQAFYSSYDLIHWLSTLSTSG 273
QY 230 HALGGIECPDLFEITADDGTRH--WLAASMDAYGIGLPM---TYAYWTGTWDEQEQHADD 285
DB 274 YLGLOVECEPMARVPVEGDEYKWLFIINP---CAPLGGSVVQYFVGDNWGTNEVPDD 330
QY 286 LPFQWLDGWQWYAAVTPSIDAPETKRIAIAMNNWYKAARDVPTDASLTYNGQNSIVR 345
DB 331 GQTRFVDLKGDFYASALYHSSA--NADVICGVGWSWNOY--TQAPQV---FRSAMTVAR 385
QY 346 ELRLARQPGGWYT---LLSTP--VAALTN--YVTATTLPDRTVDGSAVLPWNGRAYE- 396
DB 386 KFTLRDVQPNMTLSLQTLPLNLSLRDELFTAPVINSSSSLSGSPITLPSNTAFEF 445
QY 397 -IELDIAMDTATNVGISVGR---SPDGTR---HTNIGKYGADLYVDYRGPSSDLGY-SL 446
DB 446 NVTLINSYITEGCTTGCLGRIIIDSDDPYRLQISIVDVFDAASTLVINRAKAGMGWNSL 505
QY 447 APYSRAAADPDGARSVHRLIILVQTSVEFVNAGHTVLSQQVHFAEGDT 496
DB 506 FTSPFANDIYIG--NVTLYGIVDNGLLLELYVNGEKTNTDFFFLQAGT 553
RESULT 10
RAFD_ECOLI
ID RAFD_ECOLI STANDARD; PRT; 476 AA.
AC P16553;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DE 01-MAR-1992 (Rel. 21, Last annotation update)
DE Raffinose invertase (EC 3.2.1.26) (Invertase).
GN RAFD.
OS Escherichia coli.
OG Plasmid pRS2.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.

Db 173 -SSNQFDPKVFHREKSMOHCSE-----IARVKIQIF-GSANLKNWVLSNFSSGYG 226

QY 230 HALGGIECPDLFEITADDGTR-HWVLAASMDAYGIGLPM---TYAYTGTWGDGEFHADD 285

Db 227 NOYQ---MSRLIEVPIESDKSKWVFLAIP---GSPLGSGINQYFVGDFGQFVPPD 280

QY 286 LTPQWLDGWDAVYAAVTPSIDAPEKRLAIAWMNKKYAAVDPTDASDYGNGQNSIVR 345

Db 281 SQRFVDIGKDFYAFOTFSEV---EHGVLGLAWASNQYADQ-VPTNP---WRSSTSLAR 333

QY 346 ELRLARQPGGWYLLSTPVA-----LTYVTATTLPDRTDGSVAVLP-----WNGRAY 395

Db 334 NYTLR-----YVIQMLKLTANTDKSVLPDSINVVDKLLKKNVKNKPIKTNFKGSG 387

QY 396 EIELDIAMDATNWSVGR-----SPDGRHRTNIGKYGAD-----LYVDRG----- 437

Db 388 LFDENITF-KVLNLNVPSPGKTHPDILINSOELNSSVDSIKIGFSSOSLFIYIDRHPNVE 446

QY 438 ----PSDLACYSLAPYSRAAPIDPGARSVHLRLVDTSQVEFVNAGHTVLSQOVH 490

Db 447 FPKQFFTDKLAAY-LEPLD-----YDQDLRVFSLYGIVDKNIIELYFNDGTVMNTTF 500

QY 491 FAEG-----DTGISLYTDGGFAHFTGIVWREIGQ 519

Db 501 MGEGKYPHDIOIYVDTDEEPLFELESVIRELNK 533

RESULT 12

INVA_MAIZE STANDARD; PRT; 590 AA.

AC P49174;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Beta-fructofuranosidase, cell wall isozyme precursor (EC 3.2.1.26)

DE (Sucrose-6-phosphate hydrolase) (Invertase).

OS Zea mays (Maize)

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;

OC Panicoidae; Andropogoneae; Zea.

OX NCBI_TaxID=4577;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Black Mexican Sweet;

RA Shuker S.;

RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.

CC -! CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-fructofuranoside residues in beta-D-fructofuranosides.

CC -! SUBCELLULAR LOCATION: Cell wall.

CC -! INDUCTION: By wounding and bacterial infection.

CC -! SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U17695; AAA64487.1; -

DR MaizeDB; 113032; -

DR InterPro; IPR001362; GH_32.

DR Pfam; PF00251; Glyco_hydro_32; 1.

DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; FALSE NEG.

KW Hydrolase; Glycosidase; Glycoprotein; Cell wall; Signal.

FT SIGNAL 1 28 POTENTIAL

FT CHAIN 29 590 BETA-FRUCTOFURANOSIDASE, CELL WALL ISOZYME.

FT ACT_SITE 68 68 BY SIMILARITY.

FT CARBOHYD 190 190 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 341 341 N-LINKED (GLCNAC...) (POTENTIAL).

SQ SEQUENCE 590 AA; 65198 MW; 296455661E99B892 CRC64;

Query Match 11.0%; Score 312; DB 1; Length 590;

Best Local Similarity 25.3%; Pred. No. 3.2e-15;

Matches 150; Conservative 81; Mismatches 211; Indels 150; Gaps 34;

QY 19 LALIFGGA-----VPPAASAPGS-----LRVYVHMTTPPSGWLCDPQRPVTHGAYO 66

Db 21 LAURLAGASHVTHRSLEAEAPSVPASIVSPLLRTYGHFQPPMNWINDNAPLYYKGWTH 80

QY 67 LYLHLSDQNNPGG--WDHASTTGDGVAFTH-HGTVMPLRPD--FPVWSGAVVGTANTAG 121

Db 81 LFQYXPKGAVWGNIVWAHSVSRDLINWVALEPAIYPSIDSKYCGWGSATI-----LE 135

QY 122 FGAGAVVALATOPTDGRVYQEOYLWSTDGGFTTALP-----DPVIVNTD-----GRAA 172

Db 136 DGTPALYTGIDRAD--INVQOVL-----ALPKDASDPLLRWEKPEEYNPA 182

QY 173 TTPAETENAEWEPDKIHWDTARGENVCVIGRLR-----YAAFTSNLNDWTLRRNFDP 228

Db 183 TPAAGGINATQFHDPTTAWRHA-GHWRMLVGSVGRGAMLVYRSRDFRKTAK-----HP 238

QY 229 NH--ALGGI-ECPLDFEI-----TADGTRHWLAASMDAYGIGLPMTYAYWT-G 274

Db 239 LHSAAITGWCECPDFPVSGPGLOAGLDTAPGCK-YVLKSSLDL-----TRYDIYTG 291

QY 275 TWGG--EOPHADDLPQW-----LDWGMWYAAVTPSIDAPEKRLAIAWMNKKYAAAR 327

Db 292 SYDGGKDRYPPDPAGDYHRRRYDYG-NYASKTF--YDPVERRRVLLGWANE-----SD 344

QY 328 DYPDTASDGYNGONSIVRELRLARQPGWYLLSTPVAALTNYVTATTLPDRTVDGSAV 387

Db 345 SVPDDRAKAGWAGIHAIPRKIWL--DPTG-KQLLOWIHEVEKLRKAV-----SVDKLV 396

QY 388 LPWNG-----RAYETELDIAMDATNVTGIVS-----GRSPDGRHRTNIGKYGADLYV 434

Db 397 KCDHEFVTGIATYQADVEVSFELEAGTSLEKAEAFDPAVDDAQKLCGVKGADARG 456

QY 435 DRGP-----SDLAGYSLAP--YSRAAA---P 455

Db 457 GVGPGFLWLASADLOERTAVFRVFRDGHGPKVKVLMCTDPTKSSLSPLDYKFTFAGFVD 516

QY 456 IDPGARSVHLRLVDTSQVEFVNAGHTVLSQOVH--FAEG-DTGISLYTDG 504

Db 517 ADISSGKITLSLIDRSVVSFSGAGGKTCILSRVPSIAVGKDAHLYVFNNG 568

RESULT 13

INVA_ZYMMO STANDARD; PRT; 512 AA.

AC P35636;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE Sucrose-6-phosphate hydrolase E1 (EC 3.2.1.26) (Sucrase E1)

DE (Invertase E1).

GN INVA.

OS Zymomonas mobilis.

OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;

OC Zymomonas.

OX NCBI_TaxID=542;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN=Z6C;

RA MEDLINE=91291345; PubMed=1368686;

RA Yanase H., Fukushi H., Ueda N., Maeda Y., Toyoda A., Tomomura K.;

RT "Cloning, sequencing, and characterization of the intracellular invertase gene from Zymomonas mobilis.";

RL Agric. Biol. Chem. 55:1383-1390(1991).

CC -! CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-fructofuranoside residues in beta-D-fructofuranosides.

CC -! SUBCELLULAR LOCATION: Cytoplasmic.

CC -! SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.

CC -----


```

Db 65 SPAPPSRGVSDKTRFDVAGASHVSAWSNMLISWQTATYHFOQKWMNDPGLY 124
QY 61 THGAYOLYLHSDQNN--GPGGDHASTTGVAFTHHGTVP--LRPD-----FPVWSGS 111
Db 125 HKGMWHLFYQYNPDSAIWNITWGHAVSKDLI---HWLYLPFAMVPDQWYDINGVWTGS 180
QY 112 AVVGTANTAGFAGAVVALATQPTDGVKRYQBOYLWSTDGGFTTALPDVY---VNTD 168
Db 181 ATI-----LPDQIMMLYTGTDYD--VQVQNLAYPAN-----LSDPLLLDWKVK 224
QY 169 GRAATPAEIAENAEWRDPKIHV-DTARGEWCV-----IGRLRYAAFYSPNLRDWTLLR 223
Db 225 GNPVLVPPPGIGVKDFRDPDTAWTGPQNGWLLTIGSKIGTKGVALVYETSNTSKL-- 282
QY 224 NFDPYPHALGGI---ECPDLFEITA-----DDGTRHWVLAASM-----DAYGIG 264
Db 283 -LDGVLHVPCTGMWECVDFEYVSTKTNGLDTSYNGPGVKH-VLKASLDNNKQDHYAIG 340
QY 265 LPMYAYWTGTWDEQFHADDLPQWLDGWYAAVTPWPSIDAPETKRLAIAMNNWKY 324
Db 341 ---TYDLGRKNKTPDNPDELDCGIGLADYG-KYYASKTE--YDPKKERRVLCWIGETDS 394
QY 325 AARDVPTDASDGYNGQNSIVRELRLARQPGGWYTLTSTPVAALTNVYATITLPTDVTVD- 383
Db 395 ESADL-----QKQWASVQSIPRTVLYDKKTG--THLLQWPEEIESLRVGDPTV--KQVDL 446
QY 384 -GSAVLPWNGRAYEIELOIADMT-----ATNVGISVGRSPDGTGTRHTNIGRYG-- 429
Db 447 QPGSIELLRVDSAEILDIEASFVDKVALQGIIEADHVGFSCSTSGAASRGILGPGVI 506
QY 430 --AD-----LYVDRG-----PSDLAGYSLAPYSRAAIPDPGARS 462
Db 507 VIADQTLSELTPYFYISKAGDRAETHFCADQTRSEAPGVGKQVYGSVPVLDGEKHS 566
QY 463 VHLRIILVDTSQVEFVNAGHTVLSQOVH 490
Db 567 --MRLLVDHSIVESFAQGGRVTITSRIY 592

RESULT 15
INVL_CAPAN STANDARD; PRT; 640 AA.
AC P3761;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Acid beta-fructofuranosidase AIV-18 (EC 3.2.1.26) (Acid sucrose-6-
DE phosphate hydrolase) (Acid invertase).
OS Capsicum annuum (bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fruit;
RA Choi D., Lee K.-W., Kim S.;
RT "Isolation and characterization of acid invertase cDNA clone in Hot
RT pepper (Capsicum annuum L.) fruits.";
RL J. Plant Biol. 40:298-303(1997).
CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -I- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL; U87849; AAB48484.1; --
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Transmembrane; Glycoprotein.
FT TRANSMEM 34 54 POTENTIAL.
FT ACT_SITE 126 126 BY SIMILARITY.
FT CARBOHYD 126 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 612 612 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 640 AA; 70620 MW; D3C628B7A7E6870B CRC64;

Query Match 9.2%; Score 260.5; DB 1; Length 640;
Best Local Similarity 23.8%; Pred. No. 1.8e-11;
Matches 138; Conservative 77; Mismatches 221; Indels 143; Gaps 32;

QY 1 MTPAISRRVQLQAGAGALALIFGGAVPPAARASAP-GSLRAYHYHTPPSGWLCDPQPV 59
Db 72 LTPATPSRGVSGVSEKTFKDVSGTSQVSYTWSNAMLNMQRTAYHFOQKNNMNDPGL 131
QY 60 TTHGAYOLYLHSDQNN--GPGGDHASTTGVAFTHHGTVP--LRPD-----FPVWSG 110
Db 132 YHKGWHLFYQYNPDSAIWNITWGHAVSTDLI---HWLYLPFAMVPDQWYDINGVWTG 187
QY 111 SAVVGTANTAGFAGAVVALATQPTDGVKRYQBOYLWSTDGGFTTALPDVY---VNT 167
Db 188 SATI-----LPDGLIMMLYTGTDYD--VQVQNLAYPAN-----LSDPLLLDWKVK 231
QY 168 DGRAATPAEIAENAEWRDPKIHV-DTARGEWCV-----IGRLRYAAFYSPN--LRDWT 220
Db 232 QGNPVLVPPPGIGVKDFRDPDTAWTGPQNGWLLTIGSKVGTGIALVYETSNTSKL 291
QY 221 LRRNFDPYPHALGGI---ECPDLFEITADD-----GTRHWVLAASMDAYGIGLP 266
Db 292 L-----HAPGTGMWECVDFEYVSTLDANGLDTSYNGPGIKH-VLKASLDD-----N 337
QY 267 MTYAYWTGTW--GEQFHADLTTPQW-----LDGWDMYAAVTPWPSIDAPETKRLATA 317
Db 338 KQDHVYVIGTYDPVKNKFSPDN--PDLDCGIGLRLDYG-RYYASKTE--YDPKKORRVLWG 392
QY 318 WMNNKYYAARDVPTDASDGYNGQNSIVRELRLARQPGGWYTLTSTPVAALTNVYATITL 377
Db 393 WIGETDSESADL-----QKQWASVQSIPRTVLYDKKTG--THLLQWPEEIESLRSDPKV 446
QY 378 PDR-----VDGSAVLPWNGRAYEIELOIADMTAT-----NVGISVGRSPDGT 420
Db 447 KEVNLOPGSIELLHVDSAA-----QFDIEASFVDRVLTLEGIIEADVGYNCSTSGAA 499
QY 421 RHTNIGKYG-----AD-----LYVDRG-----PSDLAGYSLAPYSR 451
Db 500 SRGILGPGVGVVIADQTLSELTPYFYISKAGDRAETHFCADQTRSEAPGVAKQVYG- 558
QY 452 AAAPIDPGARSVHLRIILVDTSQVEFVNAGHTVLSQOVH 490
Db 559 SSVPLVDGEKHK-RMRLLYDHSIVESFAQGGRVTITSRIY 596

```

Search completed: May 6, 2003, 12:32:41
Job time : 12.4798 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:29:32 ; Search time 56.8904 Seconds
(without alignments)
1879.728 Million cell updates/sec

Title: US-09-868-328B-4
Perfect score: 2837
Sequence: 1 MTPAIRSRRAVLQAGAGALA.....LYTDGPAHFTGIVVREIGQ 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp-archaea.*
- 2: sp-bacteria.*
- 3: sp-fungi.*
- 4: sp-human.*
- 5: sp-invertebrate.*
- 6: sp-mammal.*
- 7: sp-mhc.*
- 8: sp-organelle.*
- 9: sp-phage.*
- 10: sp-plant.*
- 11: sp-rodent.*
- 12: sp-virus.*
- 13: sp-vertebrate.*
- 14: sp-unclassified.*
- 15: sp-rvirus.*
- 16: sp-bacteriap.*
- 17: sp-archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2837	100.0	521	2	Q9KJ00
2	2352	82.9	517	2	O50585
3	2333	82.2	530	2	Q9EV09
4	1047.5	36.9	620	2	Q93R69
5	944	33.3	516	16	O07003
6	880.5	31.0	514	16	O07003
7	659.5	23.2	395	2	P94469
8	562.5	19.8	512	2	Q45372
9	527.5	18.6	501	2	Q9LAL1
10	515	18.2	622	2	Q45155
11	513	18.1	534	2	Q9RBJ1
12	486.5	17.1	485	2	Q9RLU2
13	469.5	16.5	493	2	Q9APY5
14	468.5	16.5	493	2	Q93T55
15	467.5	16.5	943	2	Q44109
16	457	16.1	537	3	Q42801

17	448	15.8	537	3	Q96T03	Q96tu3 aspergillus
18	433	15.3	516	3	O74641	O74641 aspergillus
19	427	15.1	516	3	O74642	O74642 aspergillus
20	427	15.1	516	3	O94220	O94220 aspergillus
21	425	15.0	515	3	O9HFA5	O9Hfa5 penicillium
22	423	14.9	516	3	Q96W28	Q96w28 aspergillus
23	422.5	14.9	550	5	O02490	O02490 tritriclomo
24	421	14.8	516	3	Q8X217	Q8x217 aspergillus
25	416.5	14.7	533	3	O94224	O94224 pichia jadi
26	401.5	14.2	1142	16	O97J24	O97j24 clostridium
27	389.5	13.7	515	3	O00056	O00056 penicillium
28	386	13.6	750	2	O31411	O31411 bacillus sp
29	366.5	12.9	555	3	Q9UWF4	Q9uwf4 kluyveromy
30	356	12.5	477	2	O86076	O86076 escherichia
31	355	12.5	477	16	O8XC10	O8xc10 escherichia
32	354.5	12.5	609	3	Q9Y746	Q9y746 kluyveromy
33	320.5	11.3	591	10	Q9ZTL2	Q9ztl2 zea mays (m
34	313.5	11.1	453	3	O42878	O42878 schizosacch
35	304	10.7	513	5	Q9U0Y9	Q9u0y9 leishmania
36	304	10.7	670	10	Q8RVK8	Q8rvk8 lolium pere
37	301	10.6	581	5	Q9U0Y8	Q9u0y8 leishmania
38	301	10.6	812	2	Q9RBA5	Q9rba5 arthrobacte
39	297.5	10.5	581	10	Q43856	Q43856 vicia faba
40	291	10.3	695	3	Q9C214	Q9c214 neurospora
41	283.5	10.0	580	10	Q43799	Q43799 nicotiana t
42	283	10.0	650	10	O42722	O42722 daucus caro
43	283	10.0	661	10	Q94C05	Q94c05 ipomoea bat
44	279	9.8	636	10	Q94114	Q94114 prunus cera
45	279	9.8	646	10	Q8RVH4	Q8rvh4 cichorium i

ALIGNMENTS

RESULT 1

Q9KJ00 PRELIMINARY; PRT; 521 AA.
 ID Q9KJ00
 AC Q9KJ00;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Levan fructotransferase.
 GN LFTR.
 OS Arthrobacter ureafaciens.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococcaceae; Arthrobacter.
 OX NCBI_TaxID=37931;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K-2032;
 RA Song K.B., Rhee S.K., Yoo E.J.;
 RT "Nucleotide sequence of levan fructotransferase gene (lftr) from
 RT Arthrobacter ureafaciens K-2032.";
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
 DR EMBL; AF181254; AAF73829.1;
 DR InterPro: IPR001362; GH_32.
 DR Pfam: PF00251; Glyco_hydro_32; 1.
 KW Glycosidase; Hydrolase; Transferase.
 SQ SEQUENCE 521 AA; 56590 MW; 863F3FFC4622D2B3 CRC64;

Query Match 100.0%; Score 2837; DB 2; Length 521;
 Best Local Similarity 100.0%; Pred. No. 4.9e-179;
 Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTPAIRSRRAVLQAGAGALALIFGGAVPPAARASAPGSLRAVYHMTPTPSGWLCDPQRPVT 60
 Db 1 MTPAIRSRRAVLQAGAGALALIFGGAVPPAARASAPGSLRAVYHMTPTPSGWLCDPQRPVT 60
 QY 61 THGAYQLYLLHSDQNGPGWDHASTTDCGVAFTTHGTVMLPRDPFPVWSGSAVVGTA 120
 Db 61 THGAYQLYLLHSDQNGPGWDHASTTDCGVAFTTHGTVMLPRDPFPVWSGSAVVGTA 120

```
QY 121 GFCGAGVVALATOPTDGVRRYQEQYLWSTDDGGFTTALPDVPVINTDGRAATTAEIEN 180
Db 121 GFCGAGVVALATOPTDGVRRYQEQYLWSTDDGGFTTALPDVPVINTDGRAATTAEIEN 180
QY 181 AEFRRDPKIHWDATARGEWCVIGRLRYAAYTSPNLRDWTLRNFDYPNHALGGIECPDL 240
Db 181 AEFRRDPKIHWDATARGEWCVIGRLRYAAYTSPNLRDWTLRNFDYPNHALGGIECPDL 240
QY 241 FEITADGTRHWWLAASMDAYGIGLPMYAYWTGTWDEGFHADDLTPOWLDGWDWYAA 300
Db 241 FEITADGTRHWWLAASMDAYGIGLPMYAYWTGTWDEGFHADDLTPOWLDGWDWYAA 300
QY 301 VTWPSIDAPETKRLAIAMNNWKYAAARDVPTDASDGYNGQNSIVRELRLARPCGWYTL 360
Db 301 VTWPSIDAPETKRLAIAMNNWKYAAARDVPTDASDGYNGQNSIVRELRLARPCGWYTL 360
QY 361 STPVAALTNTVTTATTLPTDPTDGSVAVLPWNGRAYEIELDIANDTATNVGISVGRSPDGT 420
Db 361 STPVAALTNTVTTATTLPTDPTDGSVAVLPWNGRAYEIELDIANDTATNVGISVGRSPDGT 420
QY 421 RHTNICKYGADLVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLVLTQSVVEFVNA 480
Db 421 RHTNICKYGADLVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLVLTQSVVEFVNA 480
QY 481 GHTVLSQQVHFAGSDGTGISLYTDGGAHFTGIVVREIGQ 519
Db 481 GHTVLSQQVHFAGSDGTGISLYTDGGAHFTGIVVREIGQ 519

RESULT 2
ID O50585 PRELIMINARY; PRT; 517 AA.
AC O50585;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Levan fructotransferase.
OS Arthrobacter nicotinovorans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=29320;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98101974; PubMed=9438987;
RA Saito K., Yokota A., Tomita F.;
RT "Molecular cloning of levan fructotransferase gene from Arthrobacter
RL Nicotinovorans GS-9 and its expression in Escherichia coli.";
RL Biosci. Biotechnol. Biochem. 61:2076-2079(1997).
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; AB001984; BAA24915.1; -
DR InterPro; IPR001362; GH_32.
DR Pfam; PR00251; Glyco_hydro.32; 1.
KW Glycosidase; Hydrolase; Transferase.
SQ SEQUENCE 517 AA; 56509 MW; D3563EF9E52A1FE3 CRC64;

Query Match 82.9%; Score 2352; DB 2; Length 517;
Best Local Similarity 81.3%; Pred. No. 4.5e-147;
Matches 421; Conservative 41; Mismatches 52; Indels 4; Gaps 2;

QY 1 MTPAISRRVLOGAGAGALALIFGAPVPPAARASAPGSLRAYVHMTPPSGWLCDDPQPV 60
Db 1 MTVDISRRVLOGAGAGALALIFGAPVPPAARASAPGSLRAYVHMTPPSGWLCDDPQPV 58
QY 61 THGAYQLYLHSDQNGPGWDHASTTGDGVAFTTHGCTVMPRLRDPFVWSSAVVGTA 120
Db 59 TNGAYQLYLHSDQNGPGWDHASTTGDGVAFTTHGCTVMPRLRDPFVWSSAVVGTA 118
QY 121 GFCGAGVVALATOPTDGVRRYQEQYLWSTDDGGFTTALPDVPVINTDGRAATTAEIEN 180
Db 119 GFCGAGVVALATOPTDGVRRYQEQYLWSTDDGGFTTALPDVPVINTDGRAATTAEIEN 176
QY 181 AEFRRDPKIHWDATARGEWCVIGRLRYAAYTSPNLRDWTLRNFDYPNHALGGIECPDL 240
Db 181 AEFRRDPKIHWDATARGEWCVIGRLRYAAYTSPNLRDWTLRNFDYPNHALGGIECPDL 240
```

```
Db 177 AEFRRDPKIHWDATARGEWCVIGRLRYAAYTSPNLRDWTLRNFDYPNHALGGIECPDL 236
QY 241 FEITADGTRHWWLAASMDAYGIGLPMYAYWTGTWDEGFHADDLTPOWLDGWDWYAA 300
Db 237 FEMTAGDGRHWWVFGASMDAYSIGLPMYAYWTGSWNGTAFIADNLTPQWLDGWDWYAA 296
QY 301 VTWPSIDAPETKRLAIAMNNWKYAAARDVPTDASDGYNGQNSIVRELRLARPCGWYTL 360
Db 297 VTWPAVEAPETKRLAIAMNNWKYAAARNVPTDASDGYNGQNSITRELRLERQSGWYTL 356
QY 361 STPVAALTNTVTTATTLPTDPTDGSVAVLPWNGRAYEIELDIANDTATNVGISVGRSPDGT 420
Db 357 STPVPALSNTATSTTLPTDPTVNGSVFVLPWNGRAYEIELDISWDTAANVGVSVGRSSDGS 416
QY 421 RHTNICKYGADLVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLVLTQSVVEFVNA 480
Db 417 RHTNICKYGDELIVDRASSEQSYALAPYTRAAAPIDANARSVHLRLVLTQSVVEFVNS 476
QY 481 GHTVLSQQVHFAGSDGTGISLYTDGGAHFTGIVVREIG 518
Db 477 GHTVWSQQVHFAGSDGTGISLYADGGAHFTGIVREIFG 514

RESULT 3
Q9EVQ9
ID Q9EVQ9 PRELIMINARY; PRT; 530 AA.
AC Q9EVQ9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Levan fructotransferase.
GN LFTM.
OS Microbacterium sp. AL-210.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Microbacteriaceae; Microbacterium.
OX NCBI_TaxID=135249;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AL-210;
RA Yang S.J., Cha J.;
RT "Levan fructotransferase gene (lftm) from Microbacterium sp. AL-210.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; AF294730; AAG01554.1; -
DR InterPro; IPR001362; GH_32.
DR Pfam; PR00251; Glyco_hydro.32; 1.
KW Glycosidase; Hydrolase; Transferase.
SQ SEQUENCE 530 AA; 57458 MW; 1A7E78D19E1D0596 CRC64;

Query Match 82.2%; Score 2333; DB 2; Length 530;
Best Local Similarity 80.2%; Pred. No. 8.2e-146;
Matches 421; Conservative 36; Mismatches 56; Indels 12; Gaps 2;

QY 4 ATISRRVLOGAGAGALALIFG-----GAVPAARASAPGSLRAYVHMTPPSGWLC 53
Db 5 SFSRMTLKGTSVGLALLIGSPTLAQSSAGLRAAPAAAS--GSLRAYVHMTPPSGWLC 62
QY 54 DPORPVTTGAYQLYLHSDQNGPGWDHASTTGDGVAFTTHGCTVMPRLRDPFVWSSAV 113
Db 63 DPORPVTTGAYQLYLHSDQNGPGWDHASTTGDGVAFTTHGCTVMPRLRDPFVWSSAV 122
QY 114 VGTANTAGGAGAVVALATOPTDGVRRYQEQYLWSTDDGGFTTALPDVPVINTDGRAAT 173
Db 123 VGTANTAGGAGAVIALATOPTDGIKRYQEQYLWSTDDGGYTFALTALPDVPVINTDGRAT 182
QY 174 TPAEINAEWFRDPKTHWDATARGEWCVIGRLRYAAYTSPNLRDWTLRNFDYPNHALG 233
Db 183 TPAEINAEWFRDPKTHWDALRGWVCVIGRARIYAFSTNLRDQWTSNFDYFNHALG 242
QY 234 GIECPDLFEITADGTRHWWLAASMDAYGIGLPMYAYWTGTWDEGFHADDLTPOWLDW 293
Db 243 GIECPDLFEITADGTRHWWLAASMDAYGIGLPMYAYWTGTWNGTHFIAGNLVPOWLDW 302
```

```

QY 294 GNDWTAAYTWPSTIDAPETKRLAIAMNNWKYAAARDVPTDASDGYNGONSIVRELRLARQP 353
DB 303 GNDWTAAYTWPSTIDAPETKRLAIAMNNWKYAAARDVPTDASDGYNGONSITRELRLERQE 362
QY 354 GGYWTLISPPVAALTNVYATTTLPDRYVDSNVLFWNGRAYEIELDIANDWTATNVGISV 413
DB 363 GGYWTLISPPVAALTNVYATTTLPDRYVDSNVLFWNGRAYEIELDIANDWTATNVGISV 422
QY 414 GRSPDCTRTNIGKYGADLYVDRGSDLAGYSLAPYSRAAAPIDPCARSVHLRIILVDTQS 473
DB 423 GRSADGARHTNIGKYGADLYVDRGSDLAGYSLAPYSRAAAPIDPCARSVHLRIILVDTQS 482
QY 474 VEVFNAGHTVLSQOVHFAEGDTGSLYTDGGAFTGIVVREIG 518
DB 483 VEVFNAGHTVLSQOVHFAEGDTGSLYTDGGAFTGIVVREIG 527

RESULT 4
Q93R69
ID Q93R69 PRELIMINARY; PRT; 620 AA.
AC Q93R69;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Levanase.
OS Microbacterium laevaniformans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococciaceae; Microbacteriaceae; Microbacterium.
OX NCBI_TaxID=36807;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15953;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpe P., Tognoni A.,
RA Totsu V., Uchiyama S., Vandenbol M., Vannier F., Vassartotti A.,
RA Viari A., Wambut R., Wedler H., Wedler H., Weitzenegeer T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; Z94043; CAB08014.1; -
DR EMBL; Z99121; CAB15451.1; -
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 620 AA; 67667 MW; 82B4909328CBAC2 CRC64;

Query Match 36.9%; Score 1047.5; DB 2; Length 620;
Best Local Similarity 45.3%; Pred. No. 5.2e-61;
Matches 243; Conservative 55; Mismatches 190; Indels 49; Gaps 16;

QY 3 PAI--SRAVLOG---AGAGALALIFGGAVPAAARASAPGSLRAYVHTPTPSGWLCDPQR 57
DB 34 PALPARRALAGLALAGSAAVAG---PAAAEEDDAP-SURARYHFTVPDHWMDPQR 89
QY 58 PVTTHGAYQLYLH-SDQNGPGG-----WDHASTTDGVAFTHHGTVMV--LRPDPFVWS 109
DB 90 PVVIGVLHYLYNADYDANPRANFTGTEWRLATSDGVAFADQGVAAKPGTNAVYDLWS 149
QY 110 GSAYVGTANTAGGAGVALATQ---PTDGVV-----KYQGYLXWSTDGDTFTALPDP 162
DB 150 GSAYVDHAGTAGGAGVAVMLVTDHPTAAQKLDASGQQAQFLWVSDVGGRTFRPDGDE 209
QY 163 VLVNTDGRAATTPAEIENAEWFRDKIHWDTARGEWCVIGRLRYAAFTSPNLRDWTLR 222
DB 210 PVIPGDGRD-----FRDKPVVNDDEQRWALIAERDRVSFTYSPDLHRWT-- 256
QY 223 RNFDYPNALGIECPDLFEITADGTRHWVLAASMDAYGIGLPMYAYWTGTWDEQFH 282
DB 257 RGEYVNAVIGTIECPDLERLADGTTTHWVNGVANGATNEPATYAWTGSFGSTFV 316
QY 283 ADDLPQMLDGMWNYAAVTPSIDAPETKRLAIAMNNWKYAAARDVPTDASDGYNGONS 342
DB 317 PDVAAPQWLHGFDMYAVTWEDPAAPLERLAIAMNNWDY-AHGAPTWPRDGTGDS 375
QY 343 IVRELRLARQGGWTTLLSTPVAALTNVYATTTLPDRYVDSNVLFWNGRAYEIELDIA 402
DB 376 ITRELRLAR-AGAGYSLAPVADALQDHATHATLTALGDRVVDGFTGLAYRGDAYEITTRIS 434

```

```

QY 403 WDTATNVGISVGRSPDGTTRHTNIGKYGADLYVDRGSDLAGYSLAPYS---RAAAPIDPG 459
DB 435 RODADNVGQLRRSADGSRHADGLTRDYAYLNRATG-----RPDSKVESRPLD-D 487
QY 460 ARSVHLRIILVDTQSVEVFNAGHTVLSQOVHFAEGDTGSLYTDGGAFTGIVVRE 516
DB 488 ADTVLRIILVDTTIEVFEVGDGRYTHSSQVFAPSGQGLALYTSGGPAVFRDLRITE 544

```

```

RESULT 5
O07003
ID O07003 PRELIMINARY; PRT; 516 AA.
AC O07003;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein yveb.
GN YVEB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Denizot F.C.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpe P., Tognoni A.,
RA Totsu V., Uchiyama S., Vandenbol M., Vannier F., Vassartotti A.,
RA Viari A., Wambut R., Wedler H., Wedler H., Weitzenegeer T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; Z94043; CAB08014.1; -
DR EMBL; Z99121; CAB15451.1; -
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
KW Glycosidase; Hydrolase; Hypothetical protein; Complete proteome.

```

```
SQ SEQUENCE 516 AA; 58978 MW; EAEC526706C40762 CRC64;
Query Match 33.3%; Score 944; DB 16; Length 516;
Best Local Similarity 41.8%; Pred. No. 2.7e-54;
Matches 205; Conservative 71; Mismatches 176; Indels 38; Gaps 15;

QY 40 RAYVHTPPSGWCLDQRPVTHGAYOLYLHS--DQNGPG-GWDHASTTGDGVAFTHHGT 97
DB 44 RAYVHTPPSGWCLDQRPVTHGAYOLYLHS--DQNGPG-GWDHASTTGDGVAFTHHGT 103
QY 98 VMP--LRPDPFVWSSGAVVGTANTAGFGAGVVALATQPTDGVKRYQEOYLWSTDGGFT 155
DB 104 AIPKYNPDGDIWGSVVDKENTAGFGKALVAIVTQPSAKDK--QEOYLWSTDKGKS 162
QY 156 F-----TALPDPIVYNTDGRAATTPAEIENAEWFRDPKIHMDTARGEWCVVIGRLRYAA 209
DB 163 FKFSYGNVMPNP---GTDD-----PRDPKVIWDDQDNKWMVMAEGSKIG 205
QY 210 FYTSPNLRDWTLRNFDPYNNHALGGIECPDLFEITADDGTRHVVLAASMDAYGIGLPMTY 269
DB 206 FYSDNLKOWHTSGF-FPEQA--GMVECPDLYMNRASDGTNKNWLGASANGKPGWKPNY 263
QY 270 AYVTGWDGEPHADLTPQWLDGWDVAAVTPSIDA--PETKRLATAWNNKRYAAR 327
DB 264 AYVTGWDGEPHADLTPQWLDGWDVAAVTPSIDA--PETKRLATAWNNKRYAAR 327
QY 328 DVPTDASGYNGQNSIVRELRLARQPGWYTLTSTPVAALTNVYVATTTLPDRTVDGSAV 387
DB 323 NTPT-MKNGFNGTDSVIRELRKEQ--DGYSLVSQPIEALEQLTVSTDEIQDVNGSKT 380
QY 388 LPNGRAYEIELDIANDTATNVGISVGRSPDGTNRHTNIGKY--GADLYVDGRPSDLAGYS 445
DB 381 LSTGTGYOLDTSLSELKNAAGVRLRESEDOQRHIDVGIFAEAGGYAYYNRAATNPDKS 440
QY 446 LAPYSRAAAPIDPGARSVHLRLVLTQSVFVFNAGHTVLSQOVHPFAEGDTGISLYTDGG 505
DB 441 -NTYVESKAPYDNKRVKHLKILVDKTTIEVFVGDKTVFSNEVFPKPEDKGITLYSDG 499
QY 506 PAHFTGIVVR 515
DB 500 TASFKNITVK 509

RESULT 6
Q97180 PRELIMINARY; PRT: 514 AA.
AC Q97180;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Levanase.
GN CAC1773.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007686; AAK79738.1; -
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR Complete proteome.
SQ SEQUENCE 514 AA; 59636 MW; 50366383B2D97004 CRC64;
```

```
Query Match 31.0%; Score 880.5; DB 16; Length 514;
Best Local Similarity 40.7%; Pred. No. 4e-50;
Matches 200; Conservative 76; Mismatches 179; Indels 37; Gaps 14;

QY 40 RAYVHTPPSGWCLDQRPVTHGAYOLYLHS--DQNGPG-GWDHASTTGDGVAFTHHGT 97
DB 45 REVVHTVNNKNDQRPVFFDGEYHYLYLYNKDYPNGNGTEWROATSKDLTWKDEGV 104
QY 98 VMP--LRPDPFVWSSGAVVGTANTAGFGAGVVALATQPTDGVKRYQEOYLWSTDGGFT 155
DB 105 CIPKYTNENGDIWGSFVVDQAQNTAGFGKALVAIVTQPSASMDK--QEQFLWYSTDRGT 163
QY 156 FTALPD-PVIVTNDGRAATTPAEIENAEWFRDPKIHMDTARGEWCVVIGRLRYAATYTP 214
DB 164 FKPSYNOPIPN-----PCTKD---FRDPKIIWDFKNNKVMVLAEGTKIGFYESY 211
QY 215 NLRDWTLRNFDPYNNHALGGIECPDLFEITADDGTRHVVLAASMDAYGIGLPMTYAYWTG 274
DB 212 NLKNW--QITGDFTNIGIVECPDIFMQSDNGNKKYKILGTSANGKVSQGEPTYAYWVG 269
QY 275 TWGGEQFHADDTLPQWLDGWDVAAVTPSIDAET--KRLAIAMNNKRYAARDVPTD 332
DB 270 NYDGKKFIADISTPKWLDYGFDMYAAVTFESENQSNLKKRYALAMNNWDY--ANNTPT- 327
QY 333 ASDGYNGQNSIVRELRLARQPGWYTLTSTPVAALTNVYVATTTLPDRTVDGSAVLPWG 392
DB 328 IQNREMGDSIVREITLSKQKDNVTSLLSKPIKRIENITTSIDQFKQISVKGIKHLKVOG 387
QY 393 RAVEITELDIAWDTATNVGISVGRSPDGTNRHTNIGKY--GADLYVDGRPSDLAGYSLAP-- 448
DB 388 KAYQLDITLWVDKAKNVGIRLRESLDRKHDVIGFTEGKYSVNR-----AYTGNPK 441
QY 449 ---YSRAAAPIDPGARSVHLRLVLTQSVFVFNAGHTVLSQOVHPFAEGDTGISLYTDGG 505
DB 442 SKKYVESRAPEDINNKVHLRFVDKVSVEVFDGKITYSNEVFPKPEDKGITLYFSIKG 501
QY 506 PAHFTGIVVREI 517
DB 502 KAVFKNITIKHI 513

RESULT 7
P94469 PRELIMINARY; PRT: 395 AA.
AC P94469;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Levanase precursor (EC 3.2.1.65) (2,6-beta-D-fructan
DE fructanohydrolase).
GN SURC.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 12980;
RX MEDLINE=98007870; PubMed=9349714;
RA Li Y., Triccas J.A., Ferenci T.;
RT "A novel levansucrase-levanase gene cluster in Bacillus
RT stearothermophilus ATCC12980."
RL Biochim. Biophys. Acta 1353:203-208(1997).
CC -!- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 2,6-BETA-D-FRUCTO
CC FURANOSIDIC LINKAGES IN 2,6-BETA-D-FRUCTANS (LEVANS) CONTAINING
CC MORE THEN 3 FRUCTOSE UNITS.
CC -!- INDUCTION: IN CONTRAST WITH LEVANSUCRASE AND SUCRASE, LEVANSASE
CC SYNTHESIS IS NOT INDUCIBLE BY SUCROSE; NO INDUCER IS KNOWN FOR
CC LEVANSASE.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; U34875; AAB97112.1; -
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
```



```
Db 129 DRGRTH-----IKYEGNPFADAMTD---FRDPKVFVHKASNKWTMILAAAGNCVRI 177
QY 211 YTSNLRDWTLRNFDPYPHALGGI-ECPLDFEITADDGTRH---WVLAASM---DAYGI 263
Db 178 YHSSNLKQKFASEFCAGNEGSHGTGWECPLDFELPV-DGDRNRKWKVMVVSIGNSDEYLE 236
QY 264 GLPMTYAYTGTWGDGQF--HADDLTPQMLDWGWDWYAATWPSIDAPETKRLAIAMNN 321
Db 237 G--SRTOYFIFGDFGRFTNENDPETVLAHDGRNYAGVTWSDIPEEDGRRLFIAWMSN 294
QY 322 WKYAARDVPTDASDGYNGQNSIVRELRLARQPGGWTLLSTPVAALTNVYVATTTLPLDR- 380
Db 295 WKT-ANHTTPEI---WRSAMTIPRSLSLSTPPEG-TRLFOTPVTELESURLEKLEWKDLE 349
QY 381 -TVDSAVLPWNGRAYEIELDIAMDATNVGISVGRSPDCTRHTNIGKYGAD---LYVDR 436
Db 350 VTTNGVNLNISGDIETIAEPELNTALEFGFKVRSAD--QETIVG-YDAEQOQLFIDR 406
QY 437 GPSDLAGYSLAPYRAAAIDPGARSVHLRLVDTOSVEFVFNAGHTVLSQQVHFAEGDT 496
Db 407 TOSGVDFCEHFPCKHGVAMPNQNRQIOMHIFVDRSSVEFVGNNGELTMTDLFFPDSSST 466
QY 497 GISLYTDGGPAHFTGIWV 514
Db 467 GIEVYAKEGNVKLVLSML 484

RESULT 10
Q45155 PRELIMINARY; PRT; 622 AA.
ID Q45155
AC Q45155;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Levanase.
GN SCRL.
OS Bacteroides fragilis.
OC Bacteria; CF6 group; Bacteroidetes; Bacteroidales; Bacteroidaceae;
OC Bacteroides.
OX NCBI_TaxID=817;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BF1;
RX MEDLINE=93259952; PubMed=8491724;
RA Blatch G.L., Woods D.R.;
RT "Molecular characterization of a fructanase produced by Bacteroides fragilis BF-1."
RL J. Bacteriol. 175:3058-3066(1993).
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; M83774; AAA22924.1; -.
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 622 AA; 70287 MW; 63424EB274CEE134 CRC64;

Query Match 18.2%; Score 515; DB 2; Length 622;
Best Local Similarity 30.2%; Pred. No. 6.4e-26;
Matches 156; Conservative 77; Mismatches 195; Indels 88; Gaps 20;

QY 40 RAYVHMTTPSGWLCQPRQPVTHGAYOLYLLHSDQNN--GPGGWDHASTTDCGVAETHHCT 97
Db 133 RPLYHHHTPLYGMMNDANGLVYKDGETHLFYQYNPYGSMGNMHWGHSVSKDLVHWEH--- 189
QY 98 VMLPDPDFP-----VWSSGAVVGTANTAGFCGAGAVVALATOPTDGVKRYQBYLYWSTD 151
Db 190 ---LEPALARDTLGHIFSGSVVDANTAGYAGAIYAFYTSASD--KNGQLOCMAYSTD 244
QY 152 GGTFTTAL-PDPVYNTDGRAATTAETIENAEWFRDPKTHWDTARGEWCVCYGRURYAAF 210
Db 245 NGRTETKYEKNPVLPFDG-----LKDFRDPKVFVYAPDQKWMVVSADKEMRF 293
QY 211 YTSNLRDWTLRNED-----YPNHALGGIECPDLFEITADDGTRH--WVLAASMDAYGI 263
```

```
Db 294 YSENLEKWTYMSGWGEYGVQPSQ-----FECPDWVELPVDGNPDHKKWALLIVNVN---- 345
QY 264 GLPMTY-----AYTGTWGDGQFHADDL--TPQWLDWGDWYAATWPSIDAPETKRLA 315
Db 346 --PGYFGSATSQYFIFGDFEGKEFKVCDNKPEIVKWLWDGCKDHYATVCFSTNG---DRTIA 400
QY 316 IAWNNWKAARDVPTDASDGYNGQNSIVRELRLARQPGGWTLLSTPVAALTNVYVATT 375
Db 401 VPWMSNQY-ANIVPTPRQ---FRSANALPRELSLYTQDGDYI-MAAAPVEETKSLRKES 455
QY 376 TLPDRIVDGS-----AVLPWNGRAYEIELDIAMDATNVGISVGRSPDCTRHTNIGKYGAD 431
Db 456 EIPAFEGVDAYHVDSLSDNKGAYEIELELAAGSABEIMGLKLFNKEGENVDIYISLPEKK 515
QY 432 LYVDRGPSDLA--GYSLAPYSRAA-----APIDPCARSVHLRLV 469
Db 516 LYMDRTKSGIVDFGRDSAPHAIEAHDRRKQNSINYVDDFALGTWAPVQK-AGNYKLDIFV 574
QY 470 DTQSEVFEVFNAGHTVLSQQVHFAEGDTGISLYTDGG 505
Db 575 DKCSVEIFLNGKIAMTNLIFFPTTYPNQMSFYSRGG 610

RESULT 11
Q9RBJ1 PRELIMINARY; PRT; 534 AA.
ID Q9RBJ1
AC Q9RBJ1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Levanase precursor.
GN LSDB.
OS Acetobacter diazotrophicus (Gluconacetobacter diazotrophicus).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.
OX NCBI_TaxID=33996;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SRT4;
RX MEDLINE=96253999; PubMed=8704949;
RA Arrieta J., Hernandez L., Coego A., Suarez V., Balmori E.,
RA Menendez C., Petit-Glatron M.F., Chambert R., Selman-Housein G.;
RT "Molecular characterization of the levansucrase gene from the
RT endophytic sugarcane bacterium Acetobacter diazotrophicus SRT4."
RL Microbiology 142:1077-1085(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SRT4;
RX MEDLINE=99373266; PubMed=10441728;
RA Hernandez L., Arrieta J., Betancourt L., Falcon V., Madrazo J.,
RA Coego A., Menendez C.;
RT "Levansucrase from Acetobacter diazotrophicus SRT4 is secreted via
RT periplasm by a signal-peptide-dependent pathway."
RL Curr. Microbiol. 39:146-152(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-SRT4;
RA Menendez C., Hernandez L., Mendoza M.F., Hevia P., Selman G.,
RA Arrieta J.;
RT "Molecular cloning and expression in E. coli of an exolevanase from
RT Acetobacter diazotrophicus SRT4."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; L41732; AAF16405.1; -.
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 36 POTENTIAL.
FT CHAIN 37 534 LEVANASE.
SQ SEQUENCE 534 AA; 58367 MW; 7BDBBA4A9BA478F6 CRC64;

Query Match 18.1%; Score 513; DB 2; Length 534;
```



```
Db 133 IYTSAFKEGVSHTGTAQSLAFSTDAAGTWS-----KYAGNPVLGRGSAHFRD 180
QY 187 PKT--HMDTARGWVCVIGRLY--AAFTSPNLRDWTLLRRNFDYPNHALGGTECPDLFE 242
Db 181 PKFVEGPGAGRWIMVAVEAQHQVLYRSADLKDWYLSLTFGPNASDGEWECP--LP 238
QY 243 ITADDGTRHWLVAASMDAYGIGLPMYAYWTGTWDEQFHA--DDLTPQ----- 289
Db 239 VDGDPDNVKNVWLVNINPGAVAGSGQYFVGDFGVQFPADPSLVPTDADGNVLRHC 298
QY 290 -WLDGWDYAAVTPSIDAPETKRLAIAMNNWYAAARDVPTDASDGYNGQNSIVRELR 348
Db 299 LMDGWDYAAVTPSIDAPETKRLAIAMNNWYAAARDVPTDASDGYNGQNSIVRELR 348
QY 349 LARPGCGWTLTSTPVAALTNVYATT-----TLPDRTVDSGSAVLPWNGRAYEI 397
Db 353 LA-TVDGFPRLVORVPLDLSGEPARTIONMELTTPCCNCPTRRPPGSAQL-----I 403
QY 398 ELDIANDTATNNGISVGRSPDGRHRT--NIGKYGADLYVDRGSDIAGYSLAPYSRAAAP 455
Db 404 EAEILPCTARHIAFRLGAPDGSAAATVLSFDALTSRLTLDRRNSGNTAFTISSRRLNPQ 463
QY 456 IDPG-ARSVHLRLVDTQSVVEF 477
Db 464 VKASTGALRLKVIIDQCSVEF 486

RESULT 14
Q93755
ID Q93755 PRELIMINARY; PRT; 493 AA.
AC Q93755;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Exonulinase (EC 3.2.1.80).
GN EXO1
OS Bacillus sp. snu-7.
OC Bacteria; Firmicutes; Bacillus/clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=159251;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SNU-7;
RA Koo B.-S., Pan C.-H., Kim S.-I.;
RT "Molecular Cloning of Exonulinase Gene from Bacillus sp. snu-7 and
its Expression in Escherichia coli.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; AF366292; AAK54126.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 493 AA; 53544 MW; EC9976BD0E46 CRC64;

Query Match 16.5%; Score 468.5; DB 2; Length 493;
Best Local Similarity 29.4%; Pred. No. 5.4e-23;
Matches 148; Conservative 70; Mismatches 212; Indels 73; Gaps 18;

QY 15 GAGALALIFGAVPPAARASAGSLRAVYHMTTPSGWLCDDPQPVTHCAYQLYLHSDQ 74
Db 17 GYGLKRL--GTAASLAAATVIG--RPILHYTAKNTWLNPNGLVWVHOGVYHLFYQNPF 72
QY 75 NN--GPGGWDHASTTDGVAFTHHGTVMPLRPDPFVWMSGSAVVGTTANTAGFGAGAVVALAT 132
Db 73 DNVWGNMNGHATSTDLHLWTEHPVIAACNEEDVFGSIVVDHNGTSGFGTAEDPALVA 132
QY 133 QPTDGVK-----YQEQVLYKSTGGFTTALPDVIVNTDGRAATTPAETENAEWFRD 186
Db 133 IYTSAFKEGVSHTGTAQSLAFSTDAAGTWS-----KYAGNPVLGRGSAHFRD 180
QY 187 PKT--HMDTARGWVCVIGRLY--AAFTSPNLRDWTLLRRNFDYPNHALGGTECPDLFE 242
```

```
Db 181 PKFVEGPGAGRWIMVAVEAQHQVLYRSADLKDWYLSLTFGPNASDGEWECP--LP 238
QY 243 ITADDGTRHWLVAASMDAYGIGLPMYAYWTGTWDEQFHA--DDLTPQ----- 289
Db 239 VDGDPDNVKNVWLVNINPGAVAGSGQYFVGDFGVQFPADPSLVPTDADGNVLRHC 298
QY 290 -WLDGWDYAAVTPSIDAPETKRLAIAMNNWYAAARDVPTDASDGYNGQNSIVRELR 348
Db 299 LMDGWDYAAVTPSIDAPETKRLAIAMNNWYAAARDVPTDASDGYNGQNSIVRELR 348
QY 349 LARPGCGWTLTSTPVAALTNVYATT-----TLPDRTVDSGSAVLPWNGRAYEI 397
Db 353 LA-TVDGFPRLVORVPLDLSGEPARTIONMELTTPCCNCPTRRPPGSAQL-----I 403
QY 398 ELDIANDTATNNGISVGRSPDGRHRT--NIGKYGADLYVDRGSDIAGYSLAPYSRAAAP 455
Db 404 EAEILPCTARHIAFRLGAPDGSAAATVLSFDALTSRLTLDRRNSGNTAFTISSRRLNPQ 463
QY 456 IDPG-ARSVHLRLVDTQSVVEF 477
Db 464 VKASTGALRLKVIIDQCSVEF 486

RESULT 15
Q44109
ID Q44109 PRELIMINARY; PRT; 943 AA.
AC Q44109;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Levanase.
GN LEVJ.
OS Actinomyces naeslundii.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Actinomycetaceae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=16555;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T14V;
RX MEDLINE=95129923; PubMed=7828936;
RA Norman J.M., Bunney K.L., Giffard P.M.;
RT "Characterization of levJ, a sucrose/fructanase-encoding gene from
Actinomyces naeslundii T14V, and comparison of its product with other
sucrose-cleaving enzymes.";
RL Gene 152:93-98(1995).
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; U12274; AAA67876.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 943 AA; 99267 MW; 801A2A7D074579B9 CRC64;

Query Match 16.5%; Score 467.5; DB 2; Length 943;
Best Local Similarity 21.5%; Pred. No. 1.5e-22;
Matches 187; Conservative 83; Mismatches 22; Indels 381; Gaps 27;

QY 4 ATSRRAVLQAGAGALALIFGG---AVPPAARASAGPS-LRAVYHMTTPSGWLCDDPQRP 58
Db 19 AASTLASMPAASGAGARPVSAAPNAPAPAPAPADQOTGERWEPQSHYTPQKNWMDPNGL 78
QY 59 VTHGAYQLYLHSDQNN--GPGGWDHASTTDGVAFTHHGTVMPLRPDPFVWMSGSAVGT 116
Db 79 VYDGEYHMFYQNPPEGSDWGNMNGHAYSKDLVHMQELGVAIPHTSOYGVFGSSAVIDT 138
QY 117 ANTAGEGA---CAVALATQPTDGVKRYQEQVLYKSTGGFTTAL--PDPVIVNTDGRA 171
Db 139 KNTSLGSPDNPAWVAVMTADVGN--OSQSLAYSTDKGGTWNLYNNGDPPVL----- 189
QY 172 ATTPAEIENAEWFRDPKIHMDTARGWVCVIGRL--RYAAFTYSPNLRDWTLLRRNFDYPN 229
Db 190 -----DIGSNE-FRDPKVFWDQASGRWMTMVSVSHATEHRVSYSSPDLIHWTQSSFGGBG 243
```

QY	230	HALGGTCEDPLFEITADDGTRH--WVIA- : : :	255
Db	244	ITSAVWACPDFPLPDYDGSQYKWLVTVDASQAQYFVGSWDGTTFTPDEIPHYSBEG	303
QY	256	-----ASMDAYG- : :	262
Db	304	TTTLADFENYAGWKADCAAFSGSPATGDLPHQHGKAYVDYDFSGGDATGTLTSDFTVSS	363
QY	263	-----	262
Db	364	SYINLRTAGGKHPYNPOATGNDGGGRLLAGFDGSGWEGTWVGSFAAATPPOGATEPAQOPL	423
QY	263	-----TGLPWTY- 	271
Db	424	VNHSSAGLLNTYLDAAATGOGSDAPTGTATSPFTTIDSAY'LNLLMGNNPRPEGGADGGS	483
QY	272	-----W-----TGTW- 	276
Db	484	RVSVELLIVDGKVVRSATGRNLEELNQSWDYSDLKGKSAQIVVTDATGGNGHLLDDEV	543
QY	277	-----	276
Db	544	RASDKKASPIADNTSYNLVYDGKVVASATGNNSGTLEWTSMHVAAYKGRKARLVIEDRNG	603
QY	277	-----DGEOFHADDLTPOWLDWGDWVAATWPSIDAPETKRLAIAMWN	320
Db	604	NAEDWCHLMVDQILQSDTKAFSGADVVR-LDYKDYAAVTDWNV--PNKRRYQVGGWS	660
QY	321	NWKYAARDVPTDASDGYNGONSIVRELRLAROPGGWYTLTSTPVAAL'NYVTA-----T	374
Db	661	NWAY-VRODPTTT--WRTAMSTVREMGLTR-VNGKRLRTAQPVTALESRLTGOELIRKD	715
QY	375	TTLP-DRTVDGSVLPWNGRAVEIELDTAWDTATNVGIVSGRSPDGTHTWIG--KYGAD	431
Db	716	TDIPVGETSLGRAA---QGTSLDI SVDSLSPSSAFGLKV--LDNGEQYTLIGYDSQAKQ	770
QY	432	LYVDRGPSDLAGYSLAPYSRAAAPIDPGAR-SVHLRLILVDTQSVFVFNAGHTVLSQOVH	490
Db	771	LVVDRTHSGVTDSPKFPARSTAPLSPDSKGOVHLRIIVDAHSVEVFAADGTPVITQIVY	830
QY	491	FAGDGTGSLYTDGGPAHFTGIVVREIG	518
Db	831	PRQDATGVSLYAEGGTAHLGSLSLMHLG	858

Search completed: May 6, 2003, 12:34:43
Job time : 58.8904 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:31:02 ; Search time 19.4625 seconds
(without alignments)
784.611 Million cell updates/sec

Title: US-09-868-328B-4
Perfect score: 2837
Sequence: 1 MTPAIRRAVLQAGAGALA.....LYTDGGAHPTGIVVREIGQ 519

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	499.5	17.6	943	4	US-09-397-885-5
2	468.5	16.5	1277	4	US-09-397-885-3
3	461	16.2	923	4	US-09-397-885-1
4	348	12.3	581	4	US-09-331-581-2
5	333.5	11.8	337	4	US-09-331-581-24
6	291	10.3	332	4	US-09-331-581-23
7	275	9.7	635	1	US-08-245-809-1
8	274.5	9.7	635	1	US-08-296-624-2
9	235.5	8.3	492	4	US-09-134-001C-3895
10	205.5	7.2	630	3	US-08-860-091A-2
11	205.5	7.2	630	4	US-09-245-323A-6
12	189	6.7	615	3	US-08-860-091A-4
13	189	6.7	615	4	US-09-245-323A-8
14	166.5	5.9	626	4	US-09-019-385-2
15	147.5	5.2	565	4	US-09-142-623-11
16	141.5	5.0	635	4	US-09-142-623-1
17	140.5	5.0	574	4	US-09-142-623-13
18	119.5	4.2	3739	3	US-09-320-878-2
19	117	4.1	3491	2	US-07-642-734C-2
20	117	4.1	3491	3	US-08-439-009A-2
21	115	4.1	3739	4	US-09-105-537-33
22	115	4.1	11877	4	US-09-105-537-6
23	110.5	3.9	983	1	US-08-167-919A-10
24	110.5	3.9	983	2	US-08-449-645A-21
25	110.5	3.9	983	2	US-08-702-367A-21
26	110.5	3.9	983	3	US-08-715-106-10
27	110.5	3.9	983	5	PCT-US95-04681-21

28	106	3.7	3519	4	US-09-428-517-4	Sequence 4, Appli
29	106	3.7	4551	3	US-09-320-878-1	Sequence 1, Appli
30	106	3.7	4613	4	US-09-105-537-31	Sequence 31, Appli
31	105.5	3.7	811	1	US-08-480-604A-7	Sequence 7, Appli
32	105.5	3.7	811	2	US-08-405-496A-7	Sequence 7, Appli
33	105.5	3.7	811	4	US-08-915-136-7	Sequence 7, Appli
34	105.5	3.7	811	4	US-08-957-310-7	Sequence 7, Appli
35	105.5	3.7	812	1	US-08-480-604A-29	Sequence 29, Appli
36	105.5	3.7	812	4	US-08-915-136-29	Sequence 29, Appli
37	105.5	3.7	2710	1	US-08-480-604A-6	Sequence 6, Appli
38	105.5	3.7	2710	2	US-08-405-496A-6	Sequence 6, Appli
39	105.5	3.7	2710	4	US-08-915-136-6	Sequence 6, Appli
40	105.5	3.7	2710	4	US-08-957-310-6	Sequence 6, Appli
41	104.5	3.7	964	4	US-08-484-791-2	Sequence 2, Appli
42	104	3.7	655	4	US-09-311-626B-14	Sequence 14, Appli
43	104	3.7	3033	1	US-07-925-695-5	Sequence 5, Appli
44	103	3.6	936	5	PCT-US94-05905-22	Sequence 22, Appli
45	102.5	3.6	614	1	US-08-262-338A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-397-885-5
; Sequence 5, Application US/09397885
; Patent No. 6323007
; GENERAL INFORMATION:
; APPLICANT: Moller, Soren
; APPLICANT: Johansen, Charlotte
; APPLICANT: Schafer, Thomas
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Hoeck, Lisbeth Hedegaard
; TITLE OF INVENTION: A 2,6,-b-D-Fructan Hydrolase Enzyme And
; TITLE OF INVENTION: Process For Using The Enzyme
; FILE REFERENCE: 5540.200-US
; CURRENT APPLICATION NUMBER: US/09/397.885
; CURRENT FILING DATE: 1999-09-17
; EARLIER APPLICATION NUMBER: PA 1998 01173
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PA 1998 01623
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 60/101.615
; EARLIER FILING DATE: 1998-09-24
; EARLIER APPLICATION NUMBER: 60/111.675
; EARLIER FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; TYPE: PRT
; ORGANISM: Paenibacillus macerans
US-09-397-885-5

Query Match 17.6%; Score 499.5; DB 4; Length 943;
Best Local Similarity 30.8%; Pred. No. 9.1e-37;
Matches 157; Conservative 67; Mismatches 223; Indels 63; Gaps 19;

Qy	40	RAVYHMTPPSGWLCDPQRPVTHGAYQLYLHSDQNGPGGWDHASTTDGVAFTHHGTVM	99
Db	383	RPDYHYSFARGSADPNGLVYGEYHLF--HQD-----GGTWAHAVSTDLVHWKRLPTAL	436
Qy	100	PLRPDPFVWSSAVVGTANTAGF-----GAGAVVALATQPTDGVKRYEQEYLYWSTDGGF	154
Db	437	PWNLGLHWSGSADVADLNASGLFADSGKGLIAYTTSYNDPRPNNGRIGLAYSKDRGR	496
Qy	155	TFT-ALPDPIVINTDGRAATTPAEIENAEW-FRDPKIHNDTARGWCVCIGRLVAAYT	212
Db	497	TWEYAAERPIVIENPGKQDDP-----GGWDFRDPKVVRDEHNRRWVWVSGGDHIREFT	551
Qy	213	SPNLRDWTLRRNFDYPNALGGI-ECPDLEITADD-GTRHWVLAASMDAYGIGLPMTYA	270
Db	552	STNLIDWTLTDSFGYGVYRGVWECPDLFQALVDVDTGKWKVLMISTGAN-----PNTQG	607

```
QY 271 YWTGTDGEQFHADDLTPQW-----LDWGWDYAAVTPSIDAPETKRLAIA 317
      :: ::::
Db 608 -----SAAEYFIGELTPECKFVNDNPAGKVLATDYGKEYYASMSFAGM--PDGRRVMLA 659
QY 318 WMNWKYAARDVPTDASDGYNGQNSIVRELRLARQPGGWYTLTSTPVAALTN-----YVTA 373
      :: ::::
Db 660 WMTNWDYFPA-FPT---EGWKGLVSLPRELTQKTOKG-IRLAQTPIRELESIRGOLLFA 714
QY 374 TTTLPQRTVDG---SAVLPNGRAYEIELDIAMDTATNVGISVGRSPDGRHNTICKYGA 430
      :: ::::
Db 715 AS---DRVQADRENLLKGSSGYEIEAIEIPQASNSEFGRUREGAGKRTVVGYKT 771
QY 431 ---DLVDRGPSDLAGYSLAPYRAAIPDPGARSVHLRLILVDTQSVFVNAGHTVLSQ 487
      :: ::::
Db 772 KENEIYVDRSLSDGTGFSERFTLHQAPLPQDNRRVKRLRIFVYDSSLSLEFGDGRVFESE 831
QY 488 QVHFAEGDGTGSLYTDGGCPAHFTGIIVREI 517
Db 832 VIFPDPAHREMSLFTVGGEVNVVSLKVHAL 861

RESULT 2
US-09-397-885-3
; Sequence 3, Application US/09397885
; Patent No. 6323007
; GENERAL INFORMATION:
; APPLICANT: Moller, Soren
; APPLICANT: Johansen, Charlotte
; APPLICANT: Schafer, Thomas
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Hoeck, Lisbeth Hedegaard
; TITLE OF INVENTION: A 2,6, b-D-Fructan Hydrolase Enzyme And
; FILE REFERENCE: 5540.200-US
; CURRENT APPLICATION NUMBER: US/09/397,885
; EARLIER FILING DATE: 1999-09-17
; EARLIER APPLICATION NUMBER: PA 1998 01173
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PA 1998 01623
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 60/101,615
; EARLIER FILING DATE: 1998-09-24
; EARLIER APPLICATION NUMBER: 60/111,675
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1277
; TYPE: PRT
; ORGANISM: Paenibacillus pabuli
US-09-397-885-3

Query Match 16.5%; Score 468.5; DB 4; Length 1277;
Best Local Similarity 29.1%; Pred. No. 9.5e-34;
Matches 168; Conservative 74; Mismatches 243; Indels 93; Gaps 25;

QY 4 AISRRAVLQAGA-----GALIFGAVPPAA-----RASAPG----- 37
Db 320 ASSNRKVPSPGARRHIEIVASGLLIQVYDGYTPAAVEVTDKSYAKGNAGLVVQGMAY 379
QY 38 -----SLRAYHMTTPSGMCDPQRPVTHGAYQLYLLHSDQNGPGGW 82
Db 380 FODIYMTESMYKENYRQPHYSPLRGSASDPNGLVYVEGEYHLF--HQD---GGTWA 433
QY 83 HASTTDGVAFTHGTVMLRPLDPFVWSSGSAVVGCTANTAGF-----CAGAVVALATQPTDG 137
Db 434 HAVSSDLINWKLRLPALPNDQGHVWSSGSAIADLNNSGLFTSDSGKGLIAYTTSYHPDK 493
QY 138 VRKYQBYLWSTDGGFTFT-ALPDPVIVNTDGRAATTPAEIENAEW-FRDPKIHWDTAR 195
Db 494 PGNQRIGLAYSTDQGRNQYAKERIVIDNPKNGDDP-----GSWDFRDPKVVREDEH 548
```

```
QY 196 GEWCVCIGRLRYAAFTYSPNLRDWTLLRRNDYPNHALGGI-ECPDLFETAD-DGTRHWV 253
      :: ::::
Db 549 NRWVWVSGDHIRFETSTNLLDWTILTNFNGDYVRGGWECPDLIQLPVDGTGQRKW 608
QY 254 LAASMDAYGIGLPMYTA-----YWTG--TWDEQFHADDLTPQWL--DWGHWYAAVTPS 305
      :: ::::
Db 609 LLISTGAN-----PKTQGSDAEYEVGQLTADG-KFLNDHPAGQVLRDYGKEFYASMSFAN 663
QY 306 IDAPETKRLAIAWMNNKYAARDVPTDASDGYNGQNSIVRELRLARQPGGWYTLTSTPVA 365
      :: ::::
Db 664 M--PNQRKVMWAMTNDYDF-EFPTSS---WKGQLTIPREVSRLTDEG-VRLVQTPIT 716
QY 366 ALTNVYATTTLPDRTVDSGSAVLPNG--RAYEIELDI---AMDATNVGISVGRSPDG 419
      :: ::::
Db 717 ELQKLRHNLVSAQMTVGPKNPLGLTAGAYEIEAEVEIPANSSVTEFGQL-RQREG 775
QY 420 TRHITNICKYAD---LYVDRGPSDLAGYSLAPYRAAIPDPGARSVHLRLILVDTQSV 476
      :: ::::
Db 776 QKTT--VAYRVDQNMFDVTRTSGDVSFSDLEFTKVHEASLKPENQVKRLRIFVDESSEV 833
QY 477 FVNAGHTVLSQQVHFAEGDGTGSLYTDGGCPAHFTGIWV 514
      :: ::::
Db 834 FGNDGKVFSDVIFPDPAGRAMAFYSLGGEVKVSSMKV 871

RESULT 3
US-09-397-885-1
; Sequence 1, Application US/09397885
; Patent No. 6323007
; GENERAL INFORMATION:
; APPLICANT: Moller, Soren
; APPLICANT: Johansen, Charlotte
; APPLICANT: Schafer, Thomas
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Hoeck, Lisbeth Hedegaard
; TITLE OF INVENTION: A 2,6, b-D-Fructan Hydrolase Enzyme And
; FILE REFERENCE: 5540.200-US
; CURRENT APPLICATION NUMBER: US/09/397,885
; EARLIER FILING DATE: 1999-09-17
; EARLIER APPLICATION NUMBER: PA 1998 01173
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PA 1998 01623
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 60/101,615
; EARLIER FILING DATE: 1998-09-24
; EARLIER APPLICATION NUMBER: 60/111,675
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 923
; TYPE: PRT
; ORGANISM: Paenibacillus amylolyticus
US-09-397-885-1

Query Match 16.2%; Score 461; DB 4; Length 923;
Best Local Similarity 29.5%; Pred. No. 2.8e-33;
Matches 150; Conservative 74; Mismatches 223; Indels 62; Gaps 23;

QY 40 RAVYHMTTPSGMCDPQRPVTHGAYQLYLLHSDQNGPGGWHDHASTTQGVAFTHHGTVM 99
Db 392 RPQHYHTPIRGASDPNGLVYVEGEYHLF--HQD-----GGTWAHAVSKDMJANKRLPIAL 445
QY 100 PLRPDPFVWSSGSAVVGCTANTAGF-----CAGAVVALATQPTDGVKRYQBYLWSTDGGF 154
      :: ::::
Db 446 PWNDDHGHVWSSGSAVADMTNASGLFGDSGKGLIAYTTSFNPSPNGNQIRGLAYSXDOQR 505
QY 155 TETALPD-PVIYNTDGRAATTPAEIENAEW-FRDPKIHWDTARGWCVCVIGRLRYAAFTY 212
      :: ::::
Db 506 TWEYSKERPIVENPKSGN-----EAGNDFRDPKVRIRDDNNRWMVWVSGDHIRFET 560
QY 213 SPNLRDWTLLRRNDYPNHALGGI-ECPDLFETADDT--RHWVLAASMDAYGIGLPMY- 268
```

Db 561 STNLLDWTLTDMWGYGVGWECPDLFQLPV-DGTSSQKWMVMIETGAN----PKTG 615
Qy 269 ---YAYWTG--TWDEQFHADDLTAPOWL--DWGMDWYAAVTPSIDAPETKRLAIWANN 321
Db 616 GSDAEYFIGHTADG-KFVNNDNPACKVLRTDEGREFYASMSFANM--PDHRTVMWMTN 672
Qy 322 WKYAADVPTDASDGYNGNSIVRELRLAROPGGWYLLSTFPVAAALTYVTATTPLPORT 381
Db 673 WDYPPA-PPT---SNWKGELTIPREVSIVTTEDG-IRMVQSPIKEL-----ESLRKPLYS 722
Qy 382 VDSAVLPNGR-----AYETLIDI---AWDTATNVGIVSGRSPDGRHNTNIGKYCA 430
Db 723 ASNKSVPSSNLKGIISGAYEAEIEIPETSVTEFENIRRG--ANKQTVVGYKAS 780
Qy 431 D--LYVDGSPDLGAYSLAPYSRAAAPIDPGARSVHLRLVDTOSVEFVNAAGHTVLSQ 488
Db 781 DSRMFVDTASGETDFSNLFSKKHEAPQOMENNRKIMLLVDESSVEAFNGDKGVVFSOV 840
Qy 489 VHFAGDGTGSLYTDGGPAHFTGIVVREI 517
Db 841 IFPPASRAMSFYKGVGNVNVSLKVHQL 869

RESULT 4

US-09-331-581-2

; Sequence 2, Application US/09331581

; Patent No. 6130070

; GENERAL INFORMATION:

; APPLICANT: TOHDA, Hideki

; APPLICANT: HAMA, Yuko

; APPLICANT: KUMAGAI, Hiromichi

; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE

; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS

; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF

; FILE REFERENCE: 0059-1142-0PCT

; CURRENT APPLICATION NUMBER: US/09/331.581

; CURRENT FILING DATE: 1999-06-30

; EARLIER APPLICATION NUMBER: PCT/JP98/04929

; EARLIER FILING DATE: 1998-10-30

; EARLIER APPLICATION NUMBER: JP 9-314608

; EARLIER FILING DATE: 1997-10-31

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 2

; LENGTH: 581

; TYPE: PRT

; ORGANISM: Schizosaccharomyces pombe

US-09-331-581-2

Query Match 12.3%; Score 348; DB 4; Length 581;
Best Local Similarity 24.9%; Pred. No. 2.7e-23;
Matches 132; Conservative 82; Mismatches 226; Indels 90; Gaps 22;

Qy 28 PPARASAPGSU-----RAVYHMTPPSGMCLCDPQRPVTHGAYQLY 69
Db 53 PPFVNTTAPNCTCLGNEXYLPSPGYNATDRPKIHFTPTSSGFMNDPGLVYTGYYHMF 112
Qy 70 LHSQD--NNGPGGDHASTTDCVAF-----THGTVMLRPLRDPFVWGSVAVGTA 117
Db 113 QYSPKTLTAGEVHMGHTYVKDLIHENYPIAIPDEHENGVLSL-----PFGSAVVDVH 167
Qy 118 NTAGFGAG-----AVVALATQPTDGVKYOBYLYWSTDGGFTTALPDPIVINTDGRA 171
Db 168 NSSGLFSNDTTPERIVLIYDHTWGVA--ERQAIAYITDGGYTK-----KY 213
Qy 172 ATTPAEIENAEWFRDPKIHWDTARGEWCYVIGRLY--AAFTYSPNLRDWTLLRRNFDPN 229
Db 214 SGNPVLIDNSLQFRDPKVIWDFDANRWVIVAMSONYGIAYSSVDLLIHWELSVFSTSG 273
Qy 230 HALGIECPDLFEITADGTRH--WYLAASMDAYGTLGPM---TYAYWTGTWGDGEFHADD 285
Db 274 YLGQYECPGMARVPEGTDEYKKWLFISINP---GAPLGGSVVQYFVGDWNGTNEFVDP 330

Qy 286 LTPOWLDMGWYAAVTPSIDAPETKRLAIWANNMKYAAARDVPTDASDGYNGNSIVR 345
Db 331 GQTRVLDLGKDFYASALYHSSA--NADVIGVGWASNWQY--TNAAPTQV---FRSAMTVAR 385
Qy 346 ELRLARQGGWYT---LLSTP--VAALTN--YVTATTTLPDRTVDSGAVLPWNGRAYE- 396
Db 386 KFTLRDVPQNPMTNLTSLIQTPLNVSLLRDETLFTTAPVINSSSLSGSPITLPSNTAFEF 445
Qy 397 -IELDIAMDATNVGISVGR---SPDQTR---HTNIGKYGADLYVDGRPSDLAGY--SL 446
Db 446 NVTLISINTEGCTGYCLGRILIIISDDPYRLQSIISVDVDFAASTLVINRAQMGWNSL 505
Qy 447 APYSRAAAPIDPGARSVHLRLVDTOSVEFVFNAGHTVLSQOVHPAEGDT 496
Db 506 FTPSFANDIYIG--NVTLYGIVDNGLLLYVNGEKTYTNDFFFLQAT 553

RESULT 5

US-09-331-581-24

; Sequence 24, Application US/09331581

; Patent No. 6130070

; GENERAL INFORMATION:

; APPLICANT: TOHDA, Hideki

; APPLICANT: HAMA, Yuko

; APPLICANT: KUMAGAI, Hiromichi

; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE

; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS

; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF

; FILE REFERENCE: 0059-1142-0PCT

; CURRENT APPLICATION NUMBER: US/09/331.581

; CURRENT FILING DATE: 1999-06-30

; EARLIER APPLICATION NUMBER: PCT/JP98/04929

; EARLIER FILING DATE: 1998-10-30

; EARLIER APPLICATION NUMBER: JP 9-314608

; EARLIER FILING DATE: 1997-10-31

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 24

; LENGTH: 337

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-09-331-581-24

Query Match 11.8%; Score 333.5; DB 4; Length 337;
Best Local Similarity 30.0%; Pred. No. 2.5e-22;
Matches 105; Conservative 55; Mismatches 139; Indels 51; Gaps 16;

Qy 24 GGAVPPAARASAPGSLRAVYHMTPPSGMCLCDP-----QRPVTHGAYQLYVLSHSDQNG 77
Db 10 GFAAKMSASMTNETSDRPLVHFTPNKGMNDPGLWYDEKDAKWHYFQ--YNPNDTVMG 67
Qy 78 -PGGDHASTTDCGVAFTHHG--TVMLRPLRDPFVWGSVAVGTANTAGFCAGAV-----VA 129
Db 68 TPLFWGHATSDDLTWNEDQPIAIPKRDNSGAFSGSMVVDYNNNTSGFFNDITDPQRCVA 127
Qy 130 LATQPTDGVKYOBYLYWSTDGGFTTAL--PDPIVINTDGRAATTPAEIENAEWFRDPK 188
Db 128 IWTYNTF---ESEEQYISYTDGGYFTTEYQKNPVL-----ANSTQFRDPK 171
Qy 189 IHWDTARGEWCYVIGRLR--YAAFTYSPNLRDWTLLRRNFDPNHALGG--IDCPDLFEIT 244
Db 172 VWYEPSPKWIMTAASKSDYKIEIYSSDDLKSWKTESAF--ANEGLGYQVECPGLIEVP 229
Qy 245 A--DDGTRHVLAAASMDAYGTLGPMTYA---YWTGTWGDGEQFHADDLIPOWLDMGWNYA 299
Db 230 TEQDPSKSYWVWFISINP---GAPAGGSFNQYFVGSFNGTHFEAFDQNSRVVDFGDKYYA 286
Qy 300 AVTWPSIDAPETKRLAIWANNMKYAAARDVPTDASDGYNGNSIVRELRL 349
Db 287 LQTFENTDPTYGSA LGIAWASNWEYSA-FVPTNP---WRSSMSLVKRFSL 332

RESULT 6

US-09-331-581-23
; Sequence 23, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
; APPLICANT: TOHDA, Hideki
; APPLICANT: HAMA, Yuko
; APPLICANT: KUMAGAI, Hiromichi
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
; FILE REFERENCE: 0059-1142-OPCT
; CURRENT APPLICATION NUMBER: US/09/331,581
; CURRENT FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: PCT/JP98/04929
; EARLIER FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: JP 9-314608
; EARLIER FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Schwanniomycetes occidentalis
US-09-331-581-23

Query Match 10.3%; Score 291; DB 4; Length 332;
Best Local Similarity 28.1%; Pred. No. 1.8e-18;

Matches 94; Conservative 58; Mismatches 125; Indels 58; Gaps 17;

QY 40 RAVYHMTTPSGWLCDDPQPV--TTGAYOLYLHSDQNG---PGGDHASTDGVAFTH 94
DB 26 RLPHFTPEKGWMDPNPFGYDKTAKTHLYFYQNPATAGQPLYGHATSNDLVHWD 85
QY 95 HG-TVMPLRPDPVSGSVAVCTANTAGFAGC-----VVALATOPTDGVKRYEOXYL 147
DB 86 HEMALGPEHNDGIGSGSLVDHNNITGFFNSIDPNQRIVLY---TNMFDLOTQDIA 142
QY 148 WSTDGQFTTALPDPIVNTDGRAATTPAEIENAEWFRDPKIH-----DTARGEWC 200
DB 143 FSLDGGYTTTKYENNPIDV-----SSNQFRDPKVFVHERFKSMDHGCSE--- 187
QY 201 VIGRLRYAAFYSPNLRDWTLRNFD--YPNHALGEGIECPDLEITADGCTR-HWVLAAS 257
DB 188 -IARVKIQIF-GSANKWLVNSFSSGYGNYG---MSRLIEVPIENSCKSKWVFLA 242
QY 258 MDAYGIGLPM---TVAYTGTWDGEOFHADDLTPOWLDGWDWYAAVTPSIDAPETKRL 314
DB 243 INP---GSLPGSSINQYFVGDFDGFQFVDPDDSQTRFVDIGKDFYAFQITFSEV---EHGV 296
QY 315 ATAMNNWRYAARDVPTDASDGYNGONSIVRELRL 349
DB 297 GLAWASNMQYADQ-VPTNP---WRSSTSLARNVTL 327

RESULT 7

US-08-245-809-1
; Sequence 1, Application US/08245809
; Patent No. 5665579
; GENERAL INFORMATION:

; APPLICANT: Fitzmaurice, Leona
; APPLICANT: Mirkov, Erik
; APPLICANT: Butler, William O.
; APPLICANT: Dickinson, Craig D.
; APPLICANT: Elliott, Kathryn J.
; APPLICANT: Konno, Yoshihiro
; TITLE OF INVENTION: NOVEL INVERTASE GENE(S) AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitch, Even, Tabin & Flannery
; STREET: 135 South LaSalle Street, Suite 900
; CITY: Chicago
; STATE: IL

COUNTRY: U.S.A.
ZIP: 60603-4277
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/245,809
FILING DATE: 17-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/771,331
FILING DATE: 03-OCT-1991
APPLICATION NUMBER: US 07/660,344
FILING DATE: 22-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 51514
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-552-1311
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-245-809-1

Query Match 9.7%; Score 275; DB 1; Length 635;
Best Local Similarity 24.2%; Pred. No. 1.4e-16;

Matches 137; Conservative 76; Mismatches 236; Indels 118; Gaps 28;

QY 2 TPAISRRAVLOAGAGALALIFGGAVPPAARASAPGS-LRAYVHTMPPSGWLCDDPQPV 60
DB 65 SPAPSRGVSQGVSKTFRDVGASHVSWANMLSWORTAYHQPQKNWMDNPNGLY 124
QY 61 THGAYOLYLHSDQNN--GPGGDHASTDGVAFTHHGTVM-----LRPD-----FPWWSGS 111
DB 125 HGWYHLFYQYNPDSAINWGNITWGHAVSKDLI---HWLYLPFAMVDPQDQWYDINGVMTGS 180
QY 112 AVVGTANTAGFCAGAVVALATOPTDGVKRYQOYLWSTDGQFTTALPEPVI--VNTDG 169
DB 181 AVI-----LPDGGIMMLYTGTDDY--VQVONLAYPAN-----LSDPLLDWVKFG 224
QY 170 RAATTPAEIENAEWFRDPKIH--DTARGEWCV---IGRLRYAAFYTPNLRDWTLRN 224
DB 225 NPVLVPPPGIGVKDFRDPTTANTGFGONGWLLTIGSKIGKTGVALVYETSFKL--- 281
QY 225 FYPNHALGGI---ECPDLFEITA-----DDGTRHWLAASM-----DAYGIGL 265
DB 282 LDGVLHAVPGTCGMWECVDFVPVSKTKTNGLDTSYNGPGVKH-VLKASLDDNKQDHYAIG- 339
QY 266 PMTYAYTGTWDGEOFHADDLTPOWLDGWDWYAAVTPSIDAPETKRLAIAMNNWRYA 325
DB 340 --TYDLGNKKTDPNPELDCGIGLRLDYG--KYASKTF--YDPKERRRYLWGMIGETDSE 394
QY 326 ARDVPTDASDGYNGONSIVRELRLARQPGWVTLTSTPVAALTNVYATTTLPDRFTVD-- 383
DB 395 SADL-----QKGWASVQSIPRTVLYDKKTG--THLQWPVEEIESLRVGDPTV--KOVDLQ 446
QY 384 -GSAPLWPNNGRAYEIELDIAMDT-----ATNVGISVGRSPDGTTHNIGYK----- 429
DB 447 PGSIELLRVDSAAELDEASFEVDKVALQGIIEADHGFSCSTSGGAASRGILGPGFVIV 506
QY 430 -AD-----LYVDRG-----PSDLAGYSLAPYSRAAAPIDPCARSV 463
DB 507 IADQTLSELTPVYFYISKGADGRAETHFCADQTRSEAPGVKQVYGVSSVPVLDGKHS- 565
QY 464 HLRILVDTQSVFVFNAGHTVLSQQVH 490
DB 566 -MRLLYDHSIVESFAQGGGTVITSRIY 591

RESULT 8
US-08-296-624-2
; Sequence 2, Application US/08296624
; Patent No. 5658773
; GENERAL INFORMATION:
; APPLICANT: Bennett, Alan B.
; APPLICANT: Klann, Ellen
; TITLE OF INVENTION: Tomato Acid Invertase Gene
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth A. Weber
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,624
; FILING DATE: 26-AUG-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/770,970
; FILING DATE: 07-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 2307E-036510US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 636 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-296-624-2

Query Match 9.7%; Score 274.5; DB 1; Length 636;
Best Local Similarity 24.1%; Pred. No. 1.6e-16;
Matches 137; Conservative 76; Mismatches 236; Indels 119; Gaps 28;
Qy 2 TPAISRAVLQAGAGALALIFGAVPPAARASAPGS-LRAVYHMTPPSGWLCDPQRPVT 60
Db 65 SPAPPSRGVSGVSDKTRDVAGASHVSYANSLMSQRTAYHFQPKKNMNDPNGLPLY 124
Qy 61 THGAYQLYLHSDQNN--GPGGDHASTTDGVAFTHHGTVP--LRPD-----FPVWSGS 111
Db 125 HGWYHLYQYNPOSAINGNTWGHVASKDLI---HWLYLPEAFAMVPDQWYDINGVWTGS 180
Qy 112 AVVGTTANTAGFGAGAVVALATQPTDGVKRYQOXYLYWSTDGGFTTALPDPIV---VNTD 168
Db 181 ATI-----LPDGOIMLYTGDDTDY--VQVQNLAYPAN-----LSDPLLLLDWVKFK 224
Qy 169 GRAATTTPAEIENAEWFRDPKIHV-DTARGEWCV-----IGRLRYAAFTSPNLRDWTLLRR 223
Db 225 GNPVLVPPGIGVDRFPTTAWTGPQNGWLLTIGSKIGTKGVALVYVTSNFTSFKL-- 282
Qy 224 NFDYFNHALGFI---ECPDLFEITA-----DGTGRHWVLAASM-----DAYGIG 264
Db 283 -LDGVLHVAVPTGWMCECVDFPVSTKKTNGLTSTNGPGVKH-VLKASLDONKQDHYAIG 340
Qy 265 LPMYAYWTGWDGEQFHADDLTPOWLDGWDWYAAVTPSIDAPETKRLALAWNNWKY 324
Db 341 ---TYDLGKNKWTPDNPDLDCGIGLRUDYG-KYYASKTF--YDPKRRVRLWGWIGTDS 394

Qy 325 AARDVPTDASDGYNGQNSIVRELRLAROPGGWYTLTLLSTPVAALTNYVTATTLPDRTVD- 383
Db 395 ESADL-----QKGWASVQSIPTVLYDKKGTG--THLLQWPVEIEISLRVGDPTV--KQVDL 446
Qy 384 --GSAVLPWNGRAYEIELDIADWT-----ATNNGISVGRSPDGTGRHTNICKYG-- 429
Db 447 QPGSTIELLRVSAEELDEASFEVDKVALQGIIEADHVFSCSTGGGAASRGILGPFVGI 506
Qy 430 --AD-----LYVDRG-----PSDLAGYSLAPYSRAAAPIDPGARS 462
Db 507 VIADQTLSELTPVYFYSKAGDGRAETHFCADOTRSSEAPGVGKQVYGVSSVPVLDERKHS 566
Qy 463 VHLRILVDTQSVVEFVNAGHTVLSQQVH 490
Db 567 --MRLLVDSHIVESPAQGGRTVITSRIY 592
RESULT 9
US-09-134-001C-3895
; Sequence 3895, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3895

Query Match 8.3%; Score 235.5; DB 4; Length 492;
Best Local Similarity 23.5%; Pred. No. 3.8e-13;
Matches 123; Conservative 69; Mismatches 202; Indels 129; Gaps 24;
Qy 40 RAVYHMTPPSGWLCDPQRPVTHGAYQLYLYLHS-----DQNNPGPGGDHASTTDGVAFTHH 95
Db 35 RQTHIQPEMGLLNDPNGLIYNGHY--YISHOWFPLGAVHGLKYWFYKSKDLLHFEQ 92
Qy 96 GTVMPLRPD-----FPVWSGSAVGTANTAGFGAGAVVALATQPTDGVKRYQEQ--YLYW 148
Db 93 GTL--LKPDTRYDSHGVSYSASF-----EYQNHLYMYT 124
Qy 149 STDGGFTTALPDPIV--VNTDGRAATTPAETENAE-----WPRDPKIHWDTARGEWVC 200
Db 125 GNRDQHNRISSOMIARNKNDGKTEKPKPVITHQPGGYSYSHFDPKVF--EKNSQLYA 182
Qy 201 VIG-----RLRYAAYFTSPNLRDWTLLRRFDYFNHALGGI-----ECPDLFEIT 244
Db 183 ILGAQENEMGRLLLYRSQDVVDW-----HFEGEIKTNLTQFGYMWCECPDYFRLS 232
Qy 245 ADDGTRHWVLAASMDAYGIGLPMY--AYWTGTDGEGFHADDLTPOWLDGWDWYAAVT 302
Db 233 NKDVI--LMCPQGVAEAGDKFERNYQSGYMGICDLNFNNLFFDHESFQELDNGDFEYAPQ 290
Qy 303 WPSIDAPETKRLATAMNNWKYAADVDPTDASDGYNGQNSIVRELRLARQPGWYTLST 362
Db 291 F--VDA-DGQRLILGWMG---LPDTEPTD-KEGWAHCLTIPRVLTENG-----KLKOR 338
Qy 363 PVAAL---TNVYVATTTLTPORTVDGSA-----VLPWNGRAYEIELEIDAMDTATNVGIS 412
Db 339 PFQLEDLRTNKEAL-----GYANKPKRLHPYEGKQYEMIIDILENDA-SEIYFE 389
Qy 413 VGRSPDGTGRHTNICKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSV-----HL 465

Db	290	DLYPVSTVHTNGLDMVNGPNVKYVLKQSGDEDRHDWYAGSYDINVNDKWPDPDPENDVG	349	349
Qy	263	IGLPMTYAYVTGTWGDGEQFHAD-----DLTPQMLDGMWYAAVTWP	304	304
Db	350	IGL-----RDFGKFYASKTFYDQHKRRVLWYVGETDPQKYDLSKGHANILNIP	400	400
Qy	305	S---IDAPETKRRLAIW-----MNNWKY-AARDVPTDASDGYNGONSVIRRELRLARQP	353	353
Db	401	RTVVLDL-ETKTNLIQWPIEBETENLSKKYDEFKDV-----ELR-----P	439	439
Qy	354	GGWTTLLSTPVAALITNY-VTATITLPORTVDGSAVLFWNGRAYEIELDIANDTATNVGIS	412	412
Db	440	G---ALVPLEIGTATQLDIVATFEIDQKMLEST-----LEADVLFNCTTSEG-S	484	484
Qy	413	VGRSPDGTTRHTNICKYGADLYVDVRGSPS-----DLAG-----YSLAPYSR-----	451	451
Db	485	VARSV-----LGFPGVVVLADQRSEQLPVFYIIAKDIDGTSRTVFCADETRSSKDVS	537	537
Qy	452	-----AAAPIDPGARSVHLRLILVDTSQSVFVFNAGHTVLSQQVH	490	490
Db	538	VGKWYGVSSVPVLPGKEK-YNNRLILVDHSIVEGFAQNGRTVVTISRVY	582	582
RESULT 11				
US-09-245-323A-6				
; Sequence 6, Application US/09245323A				
; Patent No. 6365800				
; GENERAL INFORMATION:				
; APPLICANT: Calimi, Perry G.				
; TITLE OF INVENTION: Transgenic Crops Accumulating Fructose Polymers and				
; TITLE OF INVENTION: Methods for their Production				
; FILE REFERENCE: BB1082 US NA				
; CURRENT APPLICATION NUMBER: US/09/245,323A				
; CURRENT FILING DATE: 1999-02-05				
; PRIOR APPLICATION NUMBER: 60/077,727				
; PRIOR FILING DATE: 1998-03-12				
; NUMBER OF SEQ ID NOS: 8				
; SOFTWARE: MICROSOFT OFFICE 97				
; SEQ ID NO 6				
; LENGTH: 630				
; TYPE: PRT				
; ORGANISM: Helianthus tuberosus				
US-09-245-323A-6				
Query Match 7.2%; Score 205.5; DB 4; Length 630;				
Best Local Similarity 21.2%; Pred. No. 2.9e-10;				
Matches 124; Conservative 77; Mismatches 202; Indels 183; Gaps				
Qy	20	ALIFGGAVPPAARASAPGSL-----RAVYHMTPPSGWLCDPQRPVTHGAYOLYYL	70	70
Db	65	ATTFTVQLPNIDLKRVPGKLDSSAEVEWQRSTYHFQDKNFISDPDGPMTYHMGWYHLFYQ	124	124
Qy	71	HSQNN--GPGGWDHASTTDCGVAFTTHGTWMP--LRPDFVWGSVAVGTANTAFGCGA	126	126
Db	125	YNPOSAIWNITWCHSVSKMINWFH---LPAMVPDH--WYDIEGVTGTSATVLPNGQ	178	178
Qy	127	VVALATOPTGVRKYQSQYLYWSTDGQFTTALPDPIV---NTDGRAATTPEIENAEW	183	183
Db	179	IIMLYS--GNAYDLSQVQCAYAVNSS-----DPLLIENKKYEGNPVLLPPPGVGKYD	229	229
Qy	184	FRDPKIHWDTARGEWCVIG-----RLRYAAFTSPNLRDWTLLRRNFDYPNHALSGIECP	238	238
Db	230	FRDPSTLWSPGDGEYRWVMGSKNINETIGCALIYHTTNFTHFELKEEVLHVAHPHTGMWECV	289	289
Qy	239	DLFEIT-----ADGGTRH-WLAAASMDA-----YG	262	262
Db	290	DLYPVSTVHTNGLDMVNGPNVKYVLKQSGDEDRHDWYAGSYDINVNDKWPDPDPENDVG	349	349
Qy	263	IGLPMTYAYVTGTWGDGEQFHAD-----DLTPQMLDGMWYAAVTWP	304	304
Db	350	IGL-----RDFGKFYASKTFYDQHKRRVLWYVGETDPQKYDLSKGHANILNIP	400	400

Db 250 GWLVVYTT-----DYTNVELLDEPLHSVPNTDMWECVDFPVSLTND-----ALDMAAY 300
QY 262 GIGLPMYATWGTWGEHFDHDLTPOMLDGWDWYAAVTPSI-----DAPE----- 310
Db 301 GSGIKHVI---KESWEGH-----GMDWYSIGTYDAINDKWTDPNPELDVGI 343
QY 311 -----TKRLAIAMNNKYYAARDVPTDASGYNGQNSIVRELRLARQ 352
Db 344 GLRCDYGRFASKSLYDPLKKRITW-----GYVGESDSA-DQDLRS- 384
QY 353 PGGWYLLSTPVAALTNVYATTTLPDRIVDGSVLPW-----NGRAYE----- 396
Db 385 ---GWATV-----YNGRTIVILDRKT-GTHLLHWPVEVESURYNGEFKEIKLEP 431
QY 397 ---IELDIAMDTATNGVI-----SVGRSPDGTGTRHTNIGRY 428
Db 432 GSIPLDIG--TATQLDIVATFEVDQAALNATSETDDIYCTISLGAORG-----SLGPF 485
QY 429 GADLYVDRGPSDLAGYSLAPYSRAAPIDPGARSVH----- 464
Db 486 GLAVLADGSLSELT-----PVYFYIAKKADGGV-STHFTCKLRSLDYDGERVYVGGTVP 540
QY 465 -----LRLVDTQSVVEFVNAAGHTVLSQQVH 490
Db 541 VLDDEELTHRLVLDHSIVEFGAOGGRVTITSRAY 574

RESULT 14

US-09-019-385-2
; Sequence 2, Application US/09019385
; Patent No. 6147280
; GENERAL INFORMATION:
; APPLICANT: Smeekens, Josephus C.M.
; APPLICANT: Ebskamp, Michael J.M.
; APPLICANT: Geerts, Hendrikus A.M.
; APPLICANT: Weisbeek, Petrus J.
; TITLE OF INVENTION: Production of Oligosaccharides in
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson & Kindness PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,385
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,470
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NL 1000064
; FILING DATE: 05-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Broderick, Thomas F.
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: ARNO18589
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-224-0709
; TELEFAX: 206-224-0779
; TELEX: 4938023
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 626 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-019-385-2

Query Match 5.9%; Score 166.5; DB 4; Length 626;
Best Local Similarity 21.7%; Pred. No. 1e-06;
Matches 135; Conservative 76; Mismatches 233; Indels 177; Gaps 36;

QY 16 AGALALIFGAV-----PPAARASAPGS-----LRVYHMTPPSGWLCDPQR 57
Db 40 ASAMAVVVVGATLLAGLRMEQAVDEAAAGGFPWSNEMLOWQSGYHFQAKNYSDPNG 99
QY 58 PVTTHGAYLYLHSDQNNPG-GMDHASTTDGVAFTH-----HGTYMPLRPDPFVW- 108
Db 100 LMVYRGWYHMFY---QYNPVGTDMW-----DGMENGHAVSRNLVQWRTLPAMVADQWY 150
QY 109 -----SGSAVV---CTA---NTAGFGAGAV---VALATOPTDGV-----RKYOEQYLYW 148
Db 151 DILGVLSSGSMVLPNGTVIMITYTGATNASAVEVQCITATPADNDPLLRWTKHPANPVIW 210
QY 149 STDGGFTFTALPPDPVIVNTDGRAATTPAEIAENAEWFRDKPIHW-DTARGEWCVCIGRL-- 205
Db 211 SPPG-----VGTKD-----PRDPTAWYDESDETWTLLGSKDD 244
QY 206 -----RYAAFTSPNLRDWTLRNFDYPNHALGGIECDLFEI---TADCGTR--HWVL 254
Db 245 HDGHHGDIAMMYKTKDFLNYELIPGILHRVVRTGEWECIDFYPGVRRSSDNSSSEMLH-VL 303
QY 255 AASM-----DAYGIGLPMYAYWTGTWGEQFHADDLTPQWLDWG---WDW---YAAVT 302
Db 304 KASMDDERHDYISLG---TYDAAANTW-----TPIDPE-LDLGIGLYRDWKGKYYASTS 352
QY 303 WPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRLARQPGGWTLTST 362
Db 353 F--YDPKNERVLMGYGEVDSKRAVD---VKGWASIQSVPTVALDEKTR--TNLLW 404
QY 363 PVAALTNYVYATTTLPDRIVD-GSAV-LPWNGRAY-----FIELDIAWDATN---VGI 411
Db 405 PVIEETLRLNATELTDVTINTGSIHPIPLRQGTTHARHAESFHLDDASAVAALNEADVGY 464
QY 412 SVGRSPDGTGTRHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAPIDPG----- 459
Db 465 NCSSSGGAVNRGALGPFGL-LVLAAG--DRRGEQTAFFVYVSRGLDGLGHLTSCODELRS 521
QY 460 -----ARSHLRILVDTQSVVEFVNAAGHTVLSQQVHFAEG--DTGI 498
Db 522 SRAKDVTKRVIGSTVPVLDDGEALSMRVLVDHSIVOGFDMGGRTTTSRYYPMESYQEARV 581
QY 499 SLYTDGGPAHFTG--IVVREI 517
Db 582 YLFNNATGASVTAERLVVHEM 602

RESULT 15

US-09-142-623-11
; Sequence 11, Application US/09142623
; Patent No. 6337201
; GENERAL INFORMATION:
; APPLICANT: Koji YANAI et al.
; TITLE OF INVENTION: -FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF
; TITLE OF INVENTION: ISOLATING -FRUCTOFURANOSIDASE GENE, SYSTEM FOR PRODUCING
; TITLE OF INVENTION: -FRUCTOFURANOSIDASE, AND -FRUCTOFURANOSIDASE VARIANT
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible

Search completed: May 6, 2003, 12:36:35
Job time : 22.4625 secs

;; OPERATING SYSTEM: MS-DOS
;; SOFTWARE: Wordperfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/142.623
;; FILING DATE: September 10, 1998
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Lee Cheng
;; REGISTRATION NUMBER: 40,949
;; REFERENCE/DOCKET NUMBER: 98-0989*/LC(WMC)/144
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-721-8200
;; TELEFAX: 202-721-8250
;; INFORMATION FOR SEQ ID NO: 11:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 565 amino acid residues
;; TYPE: Amino acid
;; STRANDEDNESS: No. 6337201 relevant
;; TOPOLOGY: No. 6337201 relevant
;; MOLECULE TYPE: protein
;; ORIGINAL SOURCE:
;; ORGANISM: Microorganism: Penicillium roqueforti IAM7254
;; FEATURE:
;; NAME/KEY: mat peptide
;; LOCATION: 1..565
;; IDENTIFICATION METHOD: E
US-09-142-623-11

Query Match 5.2%; Score 147.5; DB 4; Length 565;
Best Local Similarity 21.8%; Pred. No. 4.8e-05;
Matches 126; Conservative 68; Mismatches 226; Indels 159; Gaps 29;
QY 26 AVPPAARASAPCSL-----RAVYHMTPPSCWLCDP-----QRPVTHHCAYQLYVYVLSHSDQNG 77
DB 11 SAPPMLSTLIANASLKTWRPRAHLPLLPNSGINDGCHYTDPKT--GLFHVGLYS-----G 64
QY 78 PGGWHDASTTDCGVAFTHTGTVNPLRPDPFVWSGSAVVGVTAN-TAGFGAGAVV--ALATQP 134
DB 65 ISG-----ATTDDLVS-----TYKDLNPD---GAPSI VAGKNDPLSVFDGSLVPSIDGMP 112
QY 135 T-----DGVRYQEQYLYKSTDDGFTTALPDPIVINTDGRAATTPAEIE 179
DB 113 TLLYTSVSYLPIHWSIPYTRGSETSLAVSDGCHNFTKL-----NOGPVPTPPEAL 165
QY 180 NAEWFRDPKIHV-----DTARGEWCVIG-----RLRYAAEYTSNLRD 218
DB 166 NTAFRDPYFQSPILDKSVNSTQGTWTYVAISGGVHGVGCPQFLYRQNDADFYWEYLQ 225
QY 219 W---TLRNFDPNHALG---GIECPDLFEITADDGTRHWLAAASMDAYGIGLP----- 266
DB 226 WKKEPLNTTWKGDWAGWGWFNFGVNFSLNAGYSEDGEIFITLGAEGSGLPIVPOVS 285
QY 267 ---MTYAYWTGTWGDGFHADDLPQWLDGWDWYAAV--TWPSIDAPETKR-----L 314
DB 286 STRDLWLTGNTVNDGVSFTKPTMAGV-LDWGVSAYAAAGKILPASSQASTKSGAPDRFI 344
QY 315 AIAWNNKHY-AARDVPTDASDYGNSIVREL-----RLARQ 352
DB 345 SYVWLTGDLFEQVGKFT- AQONWTGALLPRELNVRTISNVVDNELSRESLTWSRVARE 403
QY 353 PGGWYTLTSTPVAALTNTVYVATTT-----LPDRTVDGSAVLFPN-----GRAYEIELDIW 403
DB 404 DSGQIDLETMGISISREYYSALTSGSSFVESGKTLNAGAVFPNTSPSKFFVLTANISF 463
QY 404 DT-ATNVGISVGRSPDGRTHNNIGKY----GADLYVDRGPSDLAGYSLAPYSRAAPI-- 456
DB 464 PTSARDSGTQAGFQVLSLSLESTTIYQFSNESIIVDRSNTSAA-----ARTTAGILS 516
QY 457 --DPG-----ARSVHLRLVLDYTSQVEVFVN 479
DB 517 DNEAGRLRLFDVLRNGKEQVETLELTIVVDNSVLEVYAN 555

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:34:52 ; Search time 17.4663 Seconds
(without alignments)
2563.947 Million cell updates/sec

Title: US-09-868-328B-4
Perfect score: 2837
Sequence: 1 MTPAIRRAVLOGAGAGALA.....LYTDGPAHFTGIIVREIGQ 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328255 seqs, 86286685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA.*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB_PEP.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB_PEP.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB_PEP.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB_PEP.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB_PEP.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB_PEP.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB_PEP.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB_PEP.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB_PEP.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB_PEP.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB_PEP.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB_PEP.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB_PEP.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB_PEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	499.5	17.6	943 10	US-09-969-362-5
2	468.5	16.5	1277 10	US-09-969-362-3
3	461	16.2	923 10	US-09-969-362-1
4	221	7.8	635 9	US-10-003-392-14
5	206	7.3	630 9	US-10-003-392-16
6	205.5	7.2	630 9	US-10-003-392-18
7	202.5	7.1	433 9	US-09-738-626-6405
8	189	6.7	615 9	US-10-003-392-17
9	184	6.5	608 9	US-10-003-392-2
10	173	6.1	625 9	US-10-003-392-21
11	157	5.5	600 9	US-10-003-392-20
12	155	5.5	609 9	US-10-003-392-4
13	147.5	5.2	565 9	US-09-990-385-11
14	141.5	5.0	635 9	US-09-990-385-1
15	140.5	5.0	574 9	US-09-990-385-13
16	131.5	4.6	498 9	US-09-738-626-4737
17	130.5	4.6	495 9	US-10-003-392-8
18	126	4.4	390 9	US-10-003-392-6
19	123	4.3	471 9	US-10-003-392-10

Sequence 171, Appl
Sequence 33, Appl
Sequence 33, Appl
Sequence 33, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 1, Appl
Sequence 136, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 227, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 4, Appl
Sequence 31, Appl
Sequence 31, Appl
Sequence 31, Appl
Sequence 7, Appl
Sequence 6, Appl
Sequence 11, Appl
Sequence 688, Ap
Sequence 14, Appl
Sequence 15, Appl
Sequence 16, Appl

US-09-969-362-5
; Sequence 5, Application US/09969362
; Patent No. US20020076790A1
; GENERAL INFORMATION:
; APPLICANT: Moller, Soren
; APPLICANT: Johansen, Charlotte
; APPLICANT: Schafer, Thomas
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Hoeck, Lisbeth Hedegaard
; TITLE OF INVENTION: A 2,6,-b-D-Fructan Hydrolase Enzyme And
; FILE REFERENCE: 5540.200-US
; CURRENT APPLICATION NUMBER: US/09/969,362
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 09/397,885
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: PA 1998 01623
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: 60/101,615
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/111,675
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Paenibacillus macerans
US-09-969-362-5

ALIGNMENTS

RESULT 1

US-09-969-362-5
; Sequence 5, Application US/09969362
; Patent No. US20020076790A1
; GENERAL INFORMATION:
; APPLICANT: Moller, Soren
; APPLICANT: Johansen, Charlotte
; APPLICANT: Schafer, Thomas
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Hoeck, Lisbeth Hedegaard
; TITLE OF INVENTION: A 2,6,-b-D-Fructan Hydrolase Enzyme And
; FILE REFERENCE: 5540.200-US
; CURRENT APPLICATION NUMBER: US/09/969,362
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 09/397,885
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: PA 1998 01623
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: 60/101,615
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/111,675
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Paenibacillus macerans
US-09-969-362-5

Query Match. 17.6%; Score 499.5; DB 10; Length 943;
Best Local Similarity 30.8%; Pred. No. 6.8e-33;
Matches 157; Conservative 67; Mismatches 223; Indels 63; Gaps 19;

Oy 40 RAVYHMTPPSGWLCDPQRPVTHGAYQLYLHSDQNGGWDHASTTDDGVAFTHHGTVM 99
Db 383 RPDYHYSFARCSADPNGLVYEGYHLF--HQD-----GCTWAHAVSTDLVHWKRLPAL 436
Oy 100 PLRPDPFVWSSAVVGTANTAGF-----GAGAVVALATQPTDGVKRYQEQVLYWSTDGGF 154
Db 437 PWNDLGHVWSSAVADLNHNSGLFADSGKGLIAYTTSYNPDRPNQRIGLAYSKDRGR 496

Db 446 PWNDRHGWSSGAVADMNASLFGDSGKGLIAYTTFNPDSPNGNORIGLAYSKDQGR 505
Qy 155 TPTALPD--PVIYNTDGRAATTAEIENAEW--PRDKIHWDTARGWVCVIGRLRAAFYT 212
Db 506 TWEYSKERPIVLENGKSN-----EAGNWDPRDKVIRDDENNRVWVSGDHRFTY 560
Qy 213 SPNLKRWTLRRNFDYVNHALGI--PCDLEFETADTGT--RHWVLAASNDAYGIGLPMY- 268
Db 561 STNLDDWTLTWNWGYDVGWVGGWBCPDLPV--DGTSSQKKWMMISTGAN-----PKTG 615
Qy 269 ---YAYWTG--TWDEQFHADDLTQWL--DWGWDWYAAVTPWPSIDAPETKRLATAWNN 321
Db 616 GSDAEYFIGHLTADG--KFVNDMPAGKVLRTDFGKEFYASMSFANM--PDHRTVMAMWNTN 672
Qy 322 WKYAADVPTDASDGYNGONSIVRELRLARQPGWYTLTLLSTPVAALTNVYVATTTLPT 381
Db 673 WDYPFA--EPT---SNWKGELTIPREVSLVTTEDG--IRWQSPKEL-----ESLRKPLYS 722
Qy 382 VDGSAVLPWNGR-----AYEIELDI---AWDTATNVGISVGRSPDGTNRHTNICKYGA 430
Db 723 ASNKSVPSSGNLLKGIISGAYEIEAEIETPTSTVTEFGFNIREG--ANOKTVVGYKAS 780
Qy 431 D--LYVDRGPSDLACYSLAPYSRAAPIDPGARSVHLRLILVDTSQVEVFNAGHTVLSQO 488
Db 781 DSRMEVDRTASGETDFSNLFSKKHEAPTOMENNRKMRILVDSESSVEAFGNDGKVVFSDV 840
Qy 489 VHFABGDTGISLYTDGGPAHFTGIIVVREI 517
Db 841 ITPDPASRAMSFYKGGVGNVWVSLKVHQL 869

RESULT 4

US-10-003-392-14
; Sequence 14, Application US/10003392
; Patent No. US20020170086A1

; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Caimi, Perry G.
; APPLICANT: Stoop, Johan M.
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BBI463 US NA
; CURRENT APPLICATION NUMBER: US/10/003,392
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 14

LENGTH: 635

TYPE: PRT

ORGANISM: Parthenium argentatum Grey

US-10-003-392-14

Query Match 7.8%; Score 221; DB 9; Length 635;
Best Local Similarity 21.5%; Pred. No. 4.4e-10;
Matches 126; Conservative 82; Mismatches 207; Indels 170; Gaps 29;

Qy 12 QGAGAGALALIFGGAVPPAARASAPGSL-----RAYVHMTPPSGWCLCDPQRPVTH 62
Db 69 QGA-----ATFTTQLPRIDMKRVPGELDSGADVQWQRSAYHFQDKNVISDPDGPYHM 123
Qy 63 GAYQLYLHSDQNN--GPGWDHASTTDGVAFTTHGTVMP--LRDPFPVWSSAVVGTAN 118
Db 124 GWYHLFYQYNPESAIWGNITWGHVSVDKMINWFH----LPFAVDPH--WYDIEGVMTGS 177
Qy 119 TAGFCAGAVVALATOPTDGVKRYEQYLYWSTDGGFTETALPDPIV---NTDGRAATT 175
Db 178 ATVLPNGEIMLYT--GNAYDLSQVQCLAYAVNS-----DPLLEWKKEGPNVLLP 228
Qy 176 AEIENAEWFRDPKIHWDTARGWVCVIG-----ADDGTR--HWVLAASMDA-----YGIGLPMYAYWTGTDGEQ 230

Db 229 PPGVGKDFRDPSTLWLGPDGEYRMVMSKHNHETGICALIYHTTNTHFELNEEVLHAYP 288
Qy 231 ALGGTECPDLPEIT-----ADDGTR--HWVLAASMDA-----YGIGLPMYAYWTGT 275
Db 289 HTGMWECVDLYPVSTHTNGLDMVNGPNVKYLVKQSGDEDRHDWYAIG-----SYD 340
Qy 276 WDGEQFHADDLTPQ-----WLDWGWDWYAAVTPWPSIDAPETKRLATAWNNKYAARD 328
Db 341 WYNDKWYPPD--PENDVGIGLRYDYG--KFYASKTE-----YDQHKRRVWLW----- 383
Qy 329 VPTDASDGYNGONSIVRELRLARQPGWYTLTLLSTPVAALTNVYVATTTLPT 378
Db 384 -----GYVGEDT--PEKYDLTK--GWANILNIPRTVVLDTKTKNLNLIQWPIETEKLR 432
Qy 379 ---DRTVD-----GSVLPWNGRAVE-----IELDITAMDTATNVGTS 412
Db 433 SKYDKFVDVDELPGSLPLEIGTATQLDIVATFEVDQMMLLESTLEADVLNCTTSVG-S 491
Qy 413 VGRSPDGTNRHTNICKYGADLYVDRG-----PSDLAG-----YSLAPYSRAAAPID 457
Db 492 VGRGV-----LGPPGVVVLADAQRTQLPVYFYIAKDTDGTSTRTYFCADETRSSKQVD 544
Qy 458 PGA-----RSVHLRLILVDTSQVEVFNAGHTVLSQO 490
Db 545 VGVWYGVSSVPVLPNEKYNMRLLDVHSIVEGPAQNGRTVVTGRVY 589

RESULT 5

US-10-003-392-16
; Sequence 16, Application US/10003392
; Patent No. US20020170086A1

; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Caimi, Perry G.
; APPLICANT: Stoop, Johan M.
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BBI463 US NA
; CURRENT APPLICATION NUMBER: US/10/003,392
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16

LENGTH: 630

TYPE: PRT

ORGANISM: Helianthus sp.

US-10-003-392-16

Query Match 7.3%; Score 206; DB 9; Length 630;
Best Local Similarity 21.8%; Pred. No. 7.6e-09;
Matches 125; Conservative 81; Mismatches 204; Indels 164; Gaps 29;

Qy 28 PPAARASA-----PGSL-----RAYVHMTPPSGWCLCDPQRPVTHGAYOL 67
Db 52 PPAATFTATQLPNIDLKRVPGKLDSSAEVWQRSAYHFQDKNFISDDPGPMYHGWHL 121
Qy 68 YLHSDQNN--GPGWDHASTTDGVAFTTHGTVMP--LRDPFPVWSSAVVGTANTAGFG 123
Db 122 FYQYNPESAIWGNITWGHVSVDKMINWFH----LPFAVDPH--WYDIEGVMTGSATVLP 175
Qy 124 AGAVVALATOPTDGVKRYEQYLYWSTDGGFTETALPDPIV---NTDGRAATTAEIEN 180
Db 176 NGQIIMLYT--GNAYDLSQVQCLAYAVNS-----DPLLEWKKEGPNVLPFP 226
Qy 181 AEWFRDPKIHWDTARGWVCVIG-----RLRYAAFTYSPNLRDWTLRNFDYVPHALGI 235
Db 227 YKDFRDPSTLWLGPDGEYRMVMSKHNHETGICALIYHTTNTHFELKEEVLHAYPHTGMW 286
Qy 236 ECPDLFEIT-----ADDGTR--HWVLAASMDA-----YGIGLPMYAYWTGTDGEQ 280

Db 287 ECVDLYPVSTVHTNGLMDWNGRNVKVLKQSGDEDRHDWYAG---SYDVVNDKWYPPD 343
QY 281 FHADDLTPQWLDMGWAAVATWPSIDAPETKRLATAWNNWYAAARDVPTDASDYGNGQ 340
Db 344 PENDVGIGLYDFG-KFYASKTF---YDQHKRRVLM-----GYVGE 381
QY 341 NSIVRELRLARQPGGWYTLSTPVAAL-----TNY-----VTATTTLPDRTVD----- 383
Db 382 TD-PQKYDISK---GWANILNIPRTVVLDTKTNLIQWPIETENLRSKTYDEFKDEL 437
QY 384 --GSVLPWNGRAYE-----IELDIANDTATNVGISVGRSPDGRHTN 424
Db 438 RPSGLVPLEIGTATQLDIVATFEIDOKMLESTLEADVLFNCITTSSEG-SVARGA----- 489
QY 425 IGRYGADLYVDRGSP-----DLAG-----YSLAPYSR-----AAAPI 456
Db 490 LGPFGVVVLADARSEQLPVFFIANDIDGTSRTYFCADETRSDKSVGKWKWYGSSVPV 549
QY 457 DPGARSVHLRLVDVTOSVEVFNAGHTVLSQQVH 490
Db 550 LPGEK-YNMRLLDHSHSIVEGFAQNGRTVVTSRVY 582

RESULT 6

US-10-003-392-18
; Sequence 18, Application US/10003392
; Patent No. US20020170086A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Calimi, Perry G.
; APPLICANT: Scoop, Johan M.
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BB1463 US NA
; CURRENT APPLICATION NUMBER: US/10/003,392
; CURRENT FILING DATE: 2001-10-30
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 630
; TYPE: PRT
; ORGANISM: Helianthus tuberosus
US-10-003-392-18

Query Match 7.2%; Score 205.5; DB 9; Length 630;
Best Local Similarity 21.2%; Pred. No. 8.3e-09;
Matches 124; Conservative 77; Mismatches 202; Indels 183; Gaps 29;
QY 20 ALIFGAVPPAARASAPGSL-----RAVYHMTPPSGWGLCDPQRPVTHGAYOLYYL 70
Db 65 ATTFVTQPNIDLKRVPGKLDSSAEVWQSTYHFQDKNFISDPDGPYHMGWYHLFYQ 124
QY 71 HSDONN--GPGGWDHASTTDGVAFTHGTVMP--LRPDPVWSGSAVVGTAFTAGCAGA 126
Db 125 YNPQSAIWNITWGHVSVDKMINWFH-----LPPAMVPDH--WYDIEGVTGSAATVLPNGQ 178
QY 127 VVALATQPTDGVKRYQOYLWNSTGSGFTTALPDVIV---NTDGRAATTPAEIENAEW 183
Db 179 IIMLYS--GNAYDLSQVQCLAYAVNSS-----DPLLIWKKYEGNPVLLPPPGVGYKD 229
QY 184 FRDPKTHWDTARGWCVCVIG-----RLRYAAFTSPNLRDWTLLRRNFDPNHALGEGCP 238
Db 230 FRDPSTLWSPDGEYRWVWGSKNETIGCALIYHTTFTHFELKEEVLHAPHTGWECV 289
QY 239 DLEFIT-----ADDDGTRH-WLAAASMDA-----YCG 262
Db 290 DLYPVTSTVHTNGLMDVNDGNPNVKYLKQSGDEDRHDWYAGSYDIYNDKWYPPDENDVG 349
QY 263 IGLPNTYATWTGWDGEQFHAD-----DLTPQWLDGWGDWYAAVTWP 304

Db 350 IGL-----RYDFGKFYASKTFYDQHKRRVLMNGYVGETPOKYDLSKGWANILNIP 400
QY 305 S---IDAPETKRLATAW-----MNNWKY-AARDVPTDASDYGNGSONSIVRELRLAROP 353
Db 401 RTVVLDL-EKTNLIQWPIETENLRSKTYDEFKDV-----ELR---P 439
QY 354 GSWYTLSTPVAALTNV-VTATTTLPDRTVDGSAVLPWNGRAYEIELDIANDTATNVGIS 412
Db 440 G---ALVPLEIGTATQLDIVATFEIDOKMLEST-----LEADVLFNCITTSSEG-S 484
QY 413 VGRSPDGRHTNICKGADLYVDRGSP-----DLAG-----YSLAPYSR----- 451
Db 485 VARSV-----LGPFGVVVLADARSEQLPVFFIANDIDGTSRTYFCADETRSDKSV 537
QY 452 -----AAAPIDPGARSVHLRLVDVTOSVEVFNAGHTVLSQQVH 490
Db 538 VGKWKVYSSVPVLPGEK-YNMRLLDHSHSIVEGFAQNGRTVVTSRVY 582

RESULT 7

US-09-738-626-6405
; Sequence 6405, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 6405
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6405

Query Match 7.1%; Score 202.5; DB 9; Length 433;
Best Local Similarity 24.8%; Pred. No. 9e-09;
Matches 132; Conservative 49; Mismatches 166; Indels 185; Gaps 32;
QY 38 SLRAVYHMTPPSGWGLCDPQRPVTHGAYOLYYLHSDQNNNGPG-----GWDHAST--- 86
Db 11 SLRPAVHTVPPQGRNDPNMGVYDGTLLHYIQHD-----PGFPFAPKRI:GWAHTTPTLT 65
QY 87 -TDGVAFTHGTVMP--LRPDPF-----VWSGSAVVGTAFTAGGAVVVALATQPTDGV 138
Db 66 GPGRLQWTH---LPDALYPDASDYLDGCGYSGGAVFTDGTLLKLYTGNL-----KIDGX 115
QY 139 RKQEOYLWSTD-----GGFTFTALPDVIVNTDGRAA-TTFAEIEAENAEWFRDPKTHWD 192
Db 116 RR-ATONLVEVEDPTGLMGGIHRRSPKNPLI---DGPASGFTP-----HYRDPMLSPD 164
QY 193 TARGE-WCVIGRLR-----YAAFTSPNLRDWTLLRRNFY-----PNHALGGI- 235
Db 165 ---GDGKWKVLGNORENLITGAAYLYRSTDLNWEFSGEITFDLSDAQPGSAPDLVPGGYM 221

```

Qy 353 PGGWYTLSTPVAALTNVYVTTATTTLPDRTVDGSAVLPW-----NGRAYE-----396
      ||| :      |      |      |      |      |      |      |      |      |
Db 385 --GNATV-----YNGRTIVLDRKT-GTHLLHWPVEESLRYNGQEFKEIKLEP 431
      ||| :      |      |      |      |      |      |      |      |
Qy 397 ---TELDIANDTATNVI-----SVGRSPDCTRHTNIGKY 428
      ||| ||| :      |      |      |      |      |      |      |
Db 432 GSIIPLDIG--TATQLDIVATFEVDQAALNATSETDIYGCTTSLGAAQRG-----SLGPF 485
      ||| ||| :      |      |      |      |      |      |      |
Qy 429 GADLVYDRGPGSLAGSYLAPYSRAAAPIDPGARSVH-----464
      ||| :      |      |      |      |      |      |      |
Db 486 GLAVLADGTUSELT-----PVFYIAKKADGGV-STHFCTDKLRSSLDYDGERVVYGGTVP 540
      ||| :      |      |      |      |      |      |      |
Qy 465 -----LRILVDTQSVFVFNAGHTVLSQQVH 490
      :||| ||| :      |      |      |      |      |      |
Db 541 VLDDEELTMRLLVHDHSIVEGFAOQGRVITTSRAY 574
      :||| ||| :      |      |      |      |      |      |

RESULT 9
US-10-003-392-2
; Sequence 2, Application US/10003392
; Patent No. US20020170086A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Calimi, Perry G.
; APPLICANT: Stoop, Johan M.
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BB1463 US NA
; CURRENT APPLICATION NUMBER: US/10/003,392
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Dimorphotheca sinuata
US-10-003-392-2

```

Query Match	6.5%;	Score 184;	DB 9;	Length 608;
Best Local Similarity	20.5%;	Pred. No. 4.8e-07;		
Matches 117;	Conservative 70;	Mismatches 181;	Indels 202;	Gaps

Qy	40	RAVYHMTPPSGWLCDPQRPVTHGAYQLYLLHSDQNN--CPGQMDHASTTGDGVAETHHGT	97
Db	81	RTSEHFPQAKNFYIDPNGLPFHHGWHLFYQYPGYGVNMGNSWGHVSXKDMINWF----	136
Qy	98	VMPLRPDPFV-----WSGSAAVVGTTANTAGCAGAVVALATQTDGVRKYEQEYLWSTD	151
Db	137	-----ELPVALPTEWYDIEGLVSGSTTVLPNGQIFALYVTGNANDFSQLO-----	181
Qy	152	GGFTFTALP----DPVI--VNTDGRAAT--TPAEIENAEWFRDPKLTWDTARGEVCVIG	203
Db	182	-----CKAVPNISDPLLIEWKVYDGNPILYTPPGI--GLKDYRDPVTWTPGPGKHRMIMG	236
Qy	204	RLR-----YAAFVTPSNLRDWTLLRRNFDYFNHALGGI---ECPDLFEITADGTRHWWLA	255
Db	237	SKRNKTLGLVYHHT-----DFTNYVMSDEPLHSVPNTDMQECVDFPVPSLTNS----	287
Qy	256	ASMDAYGIGLPMYAYWYGTGWDEGEQFHADDLLTPQLWDGWDGYVAAVTW-----	304
Db	288	LDMAAYGSGIKHVI--KESWEHG-----GMDWYSIGTYDASTDKWTDPNP	330
Qy	305	SIDA-----PETERLALAWNNNKYAAARDVPTDASDGYNGQNSIVRE	346
Db	331	KLDVGIGLCDYGGFFASKSLFPDLKRRVTW-----GYVGE--SDRDP	372
Qy	347	LRLARQPGWYTLTSTLSTPVAALTYNYTATTTLPDRTVGGSAVLPW-----NGRAY-	395
Db	373	QDLSR--GWATI-----YNYARTVLDLQRT--GTHLLHWPVEETSLRSNGQFEN	418

QY 396 EIP-----LDIANDTATNVCISVGRSPDGR---HTN-----IGKY 428
||| ||| : : : : : ||| : : : : :
Db 419 EIELKPGSIPLDGSATQDIDVATEVDQDALKAISETNEEYICTKSWGAAGRSLGPF 478
||| ||| : : : : : ||| : : : : :
QY 429 GADLYYDRGSDLAGY-----SLAPYSR-----AAAPIDPGA 460
||| ||| : : : : : ||| : : : : :
Db 479 GVAVLADGLTSELTPVYFYAKKTGDSVATHCTDKLRSLDIDRERVVGVSTVPVLDG- 537
||| ||| : : : : : ||| : : : : :
QY 461 RSVHLRLVDTQSVFVNAGHTVLSQQVH 490
||| ||| : : : : : ||| : : : : :
Db 538 EELTMLLVHDHVSVEGFAQGRVTITSRY 567
||| ||| : : : : : ||| : : : : :

RESULT 10

US-10-003-392-21
; Sequence 21, Application US/10003392
; Patent No. US20020170086A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Calmi, Perry G.
; APPLICANT: Stoop, Johan M.
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BB1463 US NA
; CURRENT APPLICATION NUMBER: US/10/003,392
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
; TYPE: PRT
; LENGTH: 625
; ORGANISM: Hordeum vulgare
US-10-003-392-21

Query Match 6.1%; Score 173; DB 9; Length 625;
Best Local Similarity 21.5%; Pred. No. 4e-06;
Matches 133; Conservative 77; Mismatches 234; Indels 176; Gaps 34;

QY 16 AGALALIFGAV-----PPAARASAPGS-----LRVYHMTPPSGWLCDPQR 57
||| ||| : : : : : ||| : : : : :
Db 40 ASAMVVGVLGATLGLRMEQAVDEEAAAGFPWSNEMLQWQRSGYHFOAKNMSDPNG 99
||| ||| : : : : : ||| : : : : :
QY 58 PVTHGAYQLYLLHSQDNCPG-GWDHASTDGVAFTH-----HGTVMPLRPDPFVW- 108
||| ||| : : : : : ||| : : : : :
Db 100 LMVYRGWYHMFY---QYNPVGTDMD-----DGMWGHAVSRNLVQWRPLPIAMVADQWY 150
||| ||| : : : : : ||| : : : : :
QY 109 -----SGSAVV---GTA---NTAGFCAGAV-----VALATQPTDGV-----RKYEQYLYW 148
||| ||| : : : : : ||| : : : : :
Db 151 DILGLVSGSMTVLPNCTVIMYTGATNASAVEVQCIATPADPNDPLRRWTKHPANPVIW 210
||| ||| : : : : : ||| : : : : :
QY 149 STDGGFTTALPDPIVNTDGRAATTPAEIENAEWRDPKLIHW-DTARGEWVCVIGRL-- 205
||| ||| : : : : : ||| : : : : :
Db 211 SPGG-----VGTKD-----FRDPTAMYDESDDETWTLLSKDD 244
||| ||| : : : : : ||| : : : : :
QY 206 -----RYAIFYTSPNLRDWTLRRNDYFNHALGGTECPDLFEI---TADDGTRH--HWVL 254
||| ||| : : : : : ||| : : : : :
Db 245 HDGHHGCIAMMYKTKFLNYELIPGLILHRVVTGWECEIDFPVGRSSDNSEMLH-VL 303
||| ||| : : : : : ||| : : : : :
QY 255 AASM-----DAYGIGLPMTYAYWTGWDGEQFHADDLTPQWLDWG-----WDW---YAAVY 302
||| ||| : : : : : ||| : : : : :
Db 304 KASMDDERHDYSLG---TYDSAANTW-----TPIDPE-LDLGIGLRYDWMKGFVASTS 352
||| ||| : : : : : ||| : : : : :
QY 303 WPSIDAPEYKRLAIAMNNWYKAAARDVPTDASGYNGQNSIVRELRLARQPGWYTLTST 362
||| ||| : : : : : ||| : : : : :
Db 353 F--YDPAKNRRVLMGVYGEVDSKRAVY-----VKGWASIQSVPRVTALDEKTR--TNLLW 404
||| ||| : : : : : ||| : : : : :
QY 363 PVAALNRYVTATTLTLDRTVDSGSAVLP---WNGRAYEIELDIAWDTA-----TNVGIS 412
||| ||| : : : : : ||| : : : : :
Db 405 PVVEIETLRNLATELTDVINTGVSVHIPLRQGTQDIDIASFHLSDASAAALNEADVGN 464
||| ||| : : : : : ||| : : : : :

QY 413 VGRSPDGRTHRTNIGKYGADLYYDVRGSDLAGYSLAPYSRAAADIDPG----- 459
||| ||| : : : : : ||| : : : : :
Db 465 CSSSGGAVNARGALGPFGL-LVLAAG--DRRGQTAVFYVSRGLDGLGHTSFCQDELRS 521
||| ||| : : : : : ||| : : : : :
QY 460 -----ARSHLRLVDTQSVFVNAGHTVLSQQVHFAEG--DTGIS 499
||| ||| : : : : : ||| : : : : :
Db 522 RAKDVTKRIVGTSTVPVLDGEALSMRVLVDHSIVQGFDMGGRTTMTSRVPMESYQEARVY 581
||| ||| : : : : : ||| : : : : :
QY 500 LYTDGGGPAHFTG--IVVREI 517
||| ||| : : : : : ||| : : : : :
Db 582 LFNATGASVTAERLVVHEM 601
||| ||| : : : : : ||| : : : : :

RESULT 11

US-10-003-392-20
; Sequence 20, Application US/10003392
; Patent No. US20020170086A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Calmi, Perry G.
; APPLICANT: Stoop, Johan M.
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BB1463 US NA
; CURRENT APPLICATION NUMBER: US/10/003,392
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
; TYPE: PRT
; LENGTH: 600
; ORGANISM: Triticum aestivum
US-10-003-392-20

Query Match 5.5%; Score 157; DB 9; Length 600;
Best Local Similarity 21.2%; Pred. No. 8e-05;
Matches 125; Conservative 78; Mismatches 242; Indels 146; Gaps 30;

QY 24 GGAVPPAARASAPGS-----LRVYHMTPPSGWLCDPQRPVTHGAYQLYLYH---SD 73
||| ||| : : : : : ||| : : : : :
Db 42 GLRVDEAAAGFPWSNEMLQWQRSGYHFOAKNMSDPNGLMYNGWYHMFYINPVGT 101
||| ||| : : : : : ||| : : : : :
QY 74 QNNGCGWDHASTDGVAFTHGTVMPLRDPFVW-----SGSAVV---GTA---NTA 120
||| ||| : : : : : ||| : : : : :
Db 102 WDDGM-EWGHAVSRNLVTV---RTLPIAMVADQWYDILGLVSGSMTVLPNGTIVMIYTC 156
||| ||| : : : : : ||| : : : : :
QY 121 GFCAGAV---VALATQPTDGV-----RKYEQYLYWSTDGGFTTALPDPIVNTDGRA 172
||| ||| : : : : : ||| : : : : :
Db 157 ATNASAVEVQCIATPADPNDPLRRWTKHPANFVWSPPG-----IGTKD--- 201
||| ||| : : : : : ||| : : : : :
QY 173 TTPAEIENAEWRDPKLIHW-DTARGEWVCVIGRL-----RYAIFYTSPNLRDWTLR 223
||| ||| : : : : : ||| : : : : :
Db 202 -----FRDPTAMYDESDDETWTLLSKDDHGHGCIAMMYKTKDLNTELEIP 250
||| ||| : : : : : ||| : : : : :
QY 224 NFDYPNHALGGTECPDLFEI---TADDGTRH--VLAASMDAYGIGLPMTYAYWT-CTWDG 278
||| ||| : : : : : ||| : : : : :
Db 251 GILHRVORTGWECEIDFPVGRSSDNSEMLHVLKASMD-----DERHDYSLCTYDS 304
||| ||| : : : : : ||| : : : : :
QY 279 EQFHADDLTPQWLDWG-----WDW---YAAVTPWPSIDAPEYKRLAIAMNNWYKAAARDVPT 331
||| ||| : : : : : ||| : : : : :
Db 305 AANAWTPIDPE-LDLGIGLRYDWMKGFVASTSF--YDPAKNRRVLMGVYGEVDSKRAVY-- 359
||| ||| : : : : : ||| : : : : :
QY 332 DASDGYNGQNSIVRELRLARQPGWYTLTSTPVAALNRYVTATTLTLDRTVDSGSAVLP-- 389
||| ||| : : : : : ||| : : : : :
Db 360 --VKGWASIQSVPRVTALDEKTR--TNLLWVPEIETLRNLATELSDVTLNVTGVSVHIPL 415
||| ||| : : : : : ||| : : : : :
QY 390 -WNGRAYEIE---LDIAWDTATN---VGISVGRSPDGRTHRTNIGKYGADLYYDVRGSPDL 441
||| ||| : : : : : ||| : : : : :
Db 416 LRQGTQDIDIEATFHLSDASAAALNEADVGNCSGSGAVNARGALGPFGL-LVLAAG--DR 472
||| ||| : : : : : ||| : : : : :

```

QY 442 AGYSLAPYSRAAPIDPG-----ARSVHLRLVLD 470
      | | | | | : | | |
Db 473 RGEOTAYFYVSRGLDGLHTSFQDFLRSSRAKDVTKRIGSTVPVLGDGEAFSMRVLVD 532
      | | | | | : | | |
QY 471 TQSVVEFVNAGHTVLSQQVHFAGDGTGISLY----TDGGPAHFTGIWVREI 517
      | | | | | : | | |
Db 533 HSIVOGFAMGRTTMTSRVYPMAYQEAKEYLNNATGASVMAERLNVVHEH 583
      | | | | | : | | |

RESULT 12
US-10-003-392-4
; Sequence 4, Application US/10003392
; Patent No. US20020170086A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Caimi, Perry G.
; APPLICANT: STOOP, Johan M.
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BBI463 US NA
; CURRENT APPLICATION NUMBER: US/10/003,392
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 609
; TYPE: PRT
; ORGANISM: Parthenium argentatum Grey
US-10-003-392-4

```

```
QY      470 DTQSVEFVNAGHTVLVSOQVH 490
       I   |   |   |   :   |
Db     547 DHSVVFAQGGRITATSRVY 567

RESULT 13
US-09-990-385-11
; Sequence 11, Application US/09990385
; Publication No. US20020192771A1
; GENERAL INFORMATION:
; APPLICANT: Koji YANAI et al.
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF
; ISOLATING BETA-FRUCTOFURANOSIDASE GENE, SYSTEM FOR
; BETA-FRUCTOFURANOSIDASE, AND BETA-FRUCTOFURANOSI
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/990,385
; FILING DATE: 10-Sep-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/142,623
; FILING DATE: September 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: 2001-1611
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 amino acid residues
; TYPE: Amino acid
; TOPOLOGY: No. US20020192771A1 relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Microorganism: Penicillium roqueforti IAM7254
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 1..565
; IDENTIFICATION METHOD: E
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-990-385-11

Query Match          5.2%; Score 147.5; DB 9; Length 565;
Best Local Similarity 21.8%; Pred. No. 0.00045;
Matches 126; Conservative 68; Mismatches 226; Indels 159; Gaps 29;

QY      26 AVPPAARASAPGSL-----RAYVHMTPSGWLCDP----QRPVTHGYAYQLYYLHSDDNG 77
       |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db    11 SAPENLSTLANASLFKTRWPRAHLPLPSGNIGDCGHYTDPKT--GLFHVGWLYS-----G 64

QY      78 PGGDHDASTTDCGVAFTHGTVMRLPDPFPVWSGSAAVGTAN-TAGFCAGAVV--ALATOP 134
       |  |||  |  |||  |  |||  |  |||  |  |||  |  |||  |  |||  |  |||  |
Db    65 ISG-----ATTDDLVS-----TYKLNLNPD---GAPSIVAGGKNLDPLSVFDGSGVIPSGIDMP 112

QY      135 T-----DGVRKYIQEYLWSTDGGFTFTALPDPIVNTDGRAATTPAEIE 179
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db    113 TLLTVTSYLPILHWISIPYTRGSETQSIANSVDGGHNFKL-----NOGVIIPTPFPALE 165

QY      180 NAEMFRDPKIHW-----DTARGEWCVCVI-----RLRYAAFYTSNLRD 218
```

Db 166 NVTAFRDPYVQSPILDKSVNSTQGTWYVAISGVHGVGPCQFLYRONDADFOYWEYLQ 225
QY 219 W---TLRRNFDPYVQSPILDKSVNSTQGTWYVAISGVHGVGPCQFLYRONDADFOYWEYLQ 225
Db 226 WKEPLNTTWGKDWAGGNGFNEVGNVSLNAGVSEDEGEITILGAEGLPIVPOVS 285
QY 267 ---MTYAVWTGWGDEQFHADDLTQWLDWGDWYAAV--TWPSIDAPETKR-----L 314
Db 286 SIRDMLVNTGNVNDGVSVTKPTMAGV-LDWGVSAYAAAGKILPASSQASTKSGAPDRFI 344
QY 315 AIAWMNNKY-AARDVPTDASDGYNQNSIVREL-----RLARQ 352
Db 345 SYVWLTGDLFEQVKGFP--AQQNTGALLPRELNVRTISNVVDNELSRESLTSWRVARE 403
QY 353 PGGWYLLTSPVAALNYVTATT-----LPDRVTGSAVLPWN-----GRAYEIELDIAM 403
Db 404 DSCQIDLEWGISISRETYSALTSVSGSFVSGKTLNAGAVPNTSPSSKFFVLTANISF 463
QY 404 DT-ATNVGISVGRSPDGRHTNIGKY-----GADLYYDRGPSDLAGYSLAPYSRAAAPI-- 456
Db 464 PTSARDSGIQAGFOVLSLSLESTIYYQFSNESIIIVDRSNTSAA-----ARTTAGILS 516
QY 457 --DPG-----ARSHLRILVDVTQSVFVN 479
Db 517 DNEAGRLRLFDVLRNGKEQVETLELTIVVDNSVLEYIAN 555

RESULT 14

US-09-990-385-13
; Sequence 13, Application US/09990385
; Publication No. US20020192771A1
; GENERAL INFORMATION:
; APPLICANT: KOJI YANAI et al.
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF
; ISOLATING BETA-FRUCTOFURANOSIDASE GENE, SYSTEM FOR PRODUCING
; BETA-FRUCTOFURANOSIDASE, AND BETA-FRUCTOFURANOSIDASE VARIANTS

NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESS: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., Suite 800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
FILING DATE: 10-Sep-1998
PRIORITY APPLICATION DATA:
FILING DATE: 09/14/2003

ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: 2001-1611
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 amino acid residues
STRANDEDNESS: No. US20020192771A1 relevant
TOPOLOGY: No. US20020192771A1 relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Microorganism: Aspergillus niger ACE-2-1
(ATCC 20611)

FEATURE:
NAME/KEY: mat peptide
LOCATION: 1..635
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-990-385-1
Query Match 5.0%; Score 141.5; DB 9; Length 635;
Best Local Similarity 19.1%; Pred. No. 0.0017;
Matches 125; Conservative 83; Mismatches 245; Indels 203; Gaps 28;
QY 26 AVPPAARASAPGSL-----RAYHMTTPPSGWLCDPQRPVT--THGAYQLYLHSDQNGP 78
Db 8 APPPTLWLTLPNNTLFHVMRPAHILPAEQIGDPCAHYTDPTSTGLFHVGFHL--DGDGI 65
QY 79 GGDHAS-----TTDGVAFTHHGTVMPLRPDP--VWGSASV-VGTANTAGGAGAVVA 129
Db 66 AGATTANLATYDTSNGSF-----LIQGGKNDPVAVFDGAVIPGVNNVTPTLLYTSVSF 121
QY 130 LATQ-PTDGVKRYQEQLYWSTDDGGTF-----TALPDP----- 162
Db 122 LPHWSIPVTRSETQSLAVARDGRRFDKLDQGPVIADHPFAVDVTAFRDPVFRSAKL 181
QY 163 -VIVNTDGRAATTPAEIENA-EWFRDPKIHWDTARGEWVCVIGRLRYAAFYTSPNLRD-- 218
Db 182 DVLLSLDEEVARNETAQVQAVDGVWTEKNAPWYVAVSGVGVGPAQFLYRQNGCNASEFQ 241
QY 219 -----WTLRRNFDYPNHALGGI-----ECDFLFEITADDDG 249
Db 242 YWEYLGEWQEAATNSWGDGTWAGRWGPF---NFTGNVLFLEEGHDPQTGEVFTLGT 298
QY 250 RHVLAASMDAYGIGLPM-----YAYWTTWGDGEQFHADDLTPQ---WLDGW 295
Db 299 E-----GSGLPVIVPQVSSIHMDLWAAAGEVGVGSEQKAKVEFSPSMAGFLDWGF 347
QY 296 DWYAAV--TWPSIDAPE-----TKRLAIAMNNKYYAAROVPTDASDGYNQNSIVRE 346
Db 348 SAYAAAGKVLPAASAVSKTSKGVEDRYVSVLTDGQYEQADGFPPTAQOQGWTSLLPRE 407
QY 347 LR-----LARQPG-----GWYLLSTPVAALTNVYVATTTL 377
Db 408 LKQVTENVVDNELVREEGVSNVVGESDNTARLRLTIGITARIETKALLAN--GSVTAE 465
QY 378 PORTVDGSAVLPW-----NGRAYEIELDIAMDATN-----VGISVGRSFDGTRHT 423
Db 466 EDRTLQTAADVPAQSPSSKFFVLTAALEFPASARSPLQSGFEILASELERTAIYQFS 525
QY 424 NICKYGADLYVDGRPSDLAGYSLAPYSRAAAPIDPG-----AR 461
Db 526 N-----ESLVDRSQTS-----AAAPTNPGLDSFTESGKLRLLFDVIENGEOQVE 569
QY 462 SVHLRLVDVTQSVFVFNAGHTVLSQOVHFAEGDTGISLYTDG-GPAHFTGIIVRE 516
Db 570 TDLTVVDNNAVVEYVANGRFALSTWARSWYDNTQIRFFHNGEVEGEVQFRNVSVE 625

RESULT 15

US-09-990-385-13
; Sequence 13, Application US/09990385
; Publication No. US20020192771A1
; GENERAL INFORMATION:
; APPLICANT: KOJI YANAI et al.
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF
; ISOLATING BETA-FRUCTOFURANOSIDASE GENE, SYSTEM FOR PRODUCING
; BETA-FRUCTOFURANOSIDASE, AND BETA-FRUCTOFURANOSIDASE VARIANTS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/990.385
FILING DATE: 10-Sep-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/142.623
FILING DATE: September 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: 2001-1611
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 574 amino acid residues
TYPE: Amino acid
STRANDEDNESS: No. US20020192771A1 relevant
TOPOLOGY: No. US20020192771A1 relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Microorganism: Scopolariopsis brevicaulis IFO4843
FEATURE:
NAME/KEY: mat peptide
LOCATION: 1..574
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-990-385-13

Query Match 5.08; Score 140.5; DB:9; Length 574;
Best Local Similarity 21.98; Pred. No. 0.0018;
Matches 129; Conservative 56; Mismatches 245; Indels 159; Gaps 29;

Qy 26 AVPPAARASAPGSL-----RAYVHMTTPSGWLCDDPQPVPT--THGAYQLYLLHSDQNNNGPG 79
Db 19 SAPNLSLTLANNLSFETWRPRAHVLPPQNIQIGDCPMHYTDPETGIFHVGWLYN--GNGAS 76

Qy 80 GWDHASTDGVAFTHHGTVMPLRPDPFVWGSAAV---GTANTAGFGAGAVV--ALATQP 134
Db 77 G---ATTEDLV-----TYQDLNPD-----GAQMILPGVNDPIAVFDGAVIPSGIDGKP 122

Qy 135 TDGVRKYQEOYLW-----STDGGFTTAL-PDPVIVNTDGRAATTP--- 175
Db 123 TMNTSVSYMPISWSIAYTRGSETHSLAVSSDGGKFTKLQGVPIPPFPFGANVTSWRD 182

Qy 176 -----AEIENAEWF-----RDKIHWDTARGEWCVIGR 204
Db 183 PFLFQNPQFDSLLESENGCTWYTVISGGIHGDPSPAFLYRQHDPDFQYWEYLGPMWNEGN 242

Qy 205 LRYAFTYSPNLRDWTLRNPDYNHALGGIE---CPDLFEITADDGTRHVLAAASMDA 260
Db 243 STWGS-----GDWAGRWGYNFEVINVLGDDGYNPD-GEIATVGT-EWSEFDPIKPQ 293

Qy 261 YGIGLPWTYAVTGTWDEQFHADDLTPOWLDCWMDWYAAV--TWPSIDAPETKR----- 313
Db 294 ASDNREMLWAAGNMTLEDGDIKFTPSMAGYLDWGLSAYAAAGKELPASSKPSOKSGAPDR 353

Qy 314 -LAIAMNNNKYAAARDPTDASDCYNGNSIVREL-----RLAROPGGWYTLIS 361
Db 354 FVSYLNLITGDFEGHDFET-PQQNWTSLLLPRELVSQVTPNVVDNELARETGSW--RVG 410

Qy 362 TPVAALTYNYVT-----ATTLPDRTVDGS-----AVLPWNGRAYEIELDIATNATNVG 410
Db 411 TNDTGVLELVTLKQEIARETLAEMTSGNSTEASRNVSPPGTAQSQSDSKFFVLT-AS 469

Qy 411 ISVGRSPDGTHTNIGYAGDLYVDRGPSDLAGYSLA-----PYSRAAPIDPG--- 459
Db 470 LSF---PSSARDSL-RAGFEILSEFESTVYQFSNESIIIDRSNSAAALATDGDIT 525

Qy 460 -----ARSVHLRLVDTQSVSEFVFNAGHTVLSQOV 489
Db 526 RNEFGKMRFLFDVVEGDOERIEITLDLTIVDNSIVEVHAN-GREALSTWV 573

Search completed: May 6, 2003, 12:44:13
Job time : 20.4663 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:29:32 ; Search time 57.1096 Seconds
(without alignments)
1879.728 Million cell updates/sec

Title: US-09-868-328B-1
Perfect score: 2845
Sequence: 1 MTPAIRRAVLQAGAGALA.....TDGPAHFTGIVREIQAI 521

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mhc.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	2845	100.0	521	2	Q9KJD0
2	2355	82.8	517	2	O50585
3	2335	82.1	530	2	Q9EV09
4	1047.5	36.8	620	2	Q93R69
5	944	33.2	516	16	O07003
6	880.5	30.9	514	16	Q97I80
7	659.5	23.2	395	2	P94469
8	562.5	19.8	512	2	O45372
9	527.5	18.5	501	2	Q9LAL1
10	515	18.1	622	2	O45155
11	513	18.0	534	2	Q9RBJ1
12	486.5	17.1	485	2	O8RLU2
13	469.5	16.5	493	2	Q9APV5
14	468.5	16.5	493	2	Q93T55
15	467.5	16.4	943	2	O44109
16	457	16.1	537	3	O42801

17	448	15.7	537	3	Q96TU3
18	433	15.2	516	3	O74641
19	427	15.0	516	3	O74642
20	427	15.0	516	3	O94220
21	425	14.9	515	3	O9HFA5
22	423	14.9	516	3	Q96W28
23	422.5	14.9	550	5	O02490
24	421	14.8	516	3	O8X217
25	416.5	14.6	533	3	O94224
26	406.5	14.3	1142	16	O97J24
27	389.5	13.7	515	3	O00056
28	386	13.6	750	2	O31411
29	366.5	12.9	555	3	Q9UWF4
30	356	12.5	477	2	O86076
31	355	12.5	477	16	O8XCLO
32	354.5	12.5	609	3	Q9Y746
33	320.5	11.3	591	10	Q9ZTL2
34	313.5	11.0	453	3	O42878
35	304	10.7	513	5	Q9U0Y9
36	304	10.7	670	10	Q8RVK8
37	302	10.6	812	2	Q9RBA5
38	301	10.6	581	5	Q9U0Y8
39	297.5	10.5	581	10	Q43856
40	295	10.4	695	3	Q9C214
41	285	10.0	661	10	Q94C05
42	283.5	10.0	580	10	O43799
43	283	9.9	650	10	O42722
44	279	9.8	636	10	O94114
45	279	9.8	646	10	Q8RVH4

ALIGNMENTS

RESULT 1
Q9KJD0
ID Q9KJD0; PRELIMINARY; PRT; 521 AA.
AC Q9KJD0;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Levan fructotransferase.
GN LFPA.
OS Arthrobacter ureafaciens.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=37931;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K-2032;
RA Song K.B., Rhee S.K., Yoo E.J.;
RT "Nucleotide sequence of levan fructotransferase gene (lftA) from Arthrobacter ureafaciens K-2032.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; AF181254; AAF73829.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
KW Glycosidase; Hydrolase; Transferase.
SQ SEQUENCE 521 AA; 56590 MW; 863F3FFC4622D2B3 CRC64;

Query Match 100.0%; Score 2845; DB 2; Length 521;
Best Local Similarity 100.0%; Pred. No. 1.7e-179;
Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MTPAIRRAVLQAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPTPSGWLCDQRPVT	60
Db	1	MTPAIRRAVLQAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPTPSGWLCDQRPVT	60
Qy	61	THGAYOLYLLHSDQNGPGWDHASTTDCGVAFTHTGVTMPLRDPFVWSSGSAVGTANTA	120
Db	61	THGAYOLYLLHSDQNGPGWDHASTTDCGVAFTHTGVTMPLRDPFVWSSGSAVGTANTA	120

QY	121	GF	GAGAV	VALAT	QPTD	GV	RKY	Q	BQ	Y	L	W	S	T	D	G	G	T	T	A	L	P	D	P	V	I	N	T	D	G	R	A	A	T	T	P	A	E	I	N	180																	
Db	121	GF	GAGAV	VALAT	QPTD	GV	RKY	Q	BQ	Y	L	W	S	T	D	G	G	T	T	A	L	P	D	P	V	I	N	T	D	G	R	A	A	T	T	P	A	E	I	N	180																	
QY	181	AE	W	FR	OP	K	I	H	W	D	T	A	R	G	E	W	C	V	I	G	L	R	Y	A	F	Y	T	S	P	N	L	R	D	W	T	L	R	N	E	D	Y	PN	H	A	L	G	G	I	E	C	P	D	L	240				
Db	181	AE	W	FR	OP	K	I	H	W	D	T	A	R	G	E	W	C	V	I	G	L	R	Y	A	F	Y	T	S	P	N	L	R	D	W	T	L	R	N	E	D	Y	PN	H	A	L	G	G	I	E	C	P	D	L	240				
QY	241	FE	I	T	A	D	D	G	T	R	H	W	L	A	A	S	M	D	A	Y	G	I	G	L	P	M	T	Y	A	Y	T	G	T	W	D	G	E	Q	F	H	A	D	D	L	T	P	O	W	L	D	W	G	W	D	W	A	A	300
Db	241	FE	I	T	A	D	D	G	T	R	H	W	L	A	A	S	M	D	A	Y	G	I	G	L	P	M	T	Y	A	Y	T	G	T	W	D	G	E	Q	F	H	A	D	D	L	T	P	O	W	L	D	W	G	W	D	W	A	A	300
QY	301	V	T	W	P	S	I	D	A	P	E	T	K	L	A	I	A	M	N	N	K	Y	A	A	R	D	V	P	T	D	A	S	D	G	Y	N	G	S	N	I	V	R	E	L	R	L	A	R	O	P	G	W	Y	T	L	360		
Db	301	V	T	W	P	S	I	D	A	P	E	T	K	L	A	I	A	M	N	N	K	Y	A	A	R	D	V	P	T	D	A	S	D	G	Y	N	G	S	N	I	V	R	E	L	R	L	A	R	O	P	G	W	Y	T	L	360		
QY	361	S	T	P	V	A	L	T	N	V	T	A	T	T	L	P	O	R	T	D	V	G	S	A	V	L	P	W	N	G	R	A	Y	E	I	L	D	A	M	T	A	T	N	V	G	I	S	V	G	R	S	P	D	G	T	420		
Db	361	S	T	P	V	A	L	T	N	V	T	A	T	T	L	P	O	R	T	D	V	G	S	A	V	L	P	W	N	G	R	A	Y	E	I	L	D	A	M	T	A	T	N	V	G	I	S	V	G	R	S	P	D	G	T	420		
QY	421	R	H	T	N	I	C	K	Y	G	A	D	L	V	D	R	G	P	S	D	L	A	G	Y	S	L	A	P	I	D	P	G	A	R	S	V	H	L	R	I	L	V	D	T	O	S	V	E	F	V	N	A	480					
Db	421	R	H	T	N	I	C	K	Y	G	A	D	L	V	D	R	G	P	S	D	L	A	G	Y	S	L	A	P	I	D	P	G	A	R	S	V	H	L	R	I	L	V	D	T	O	S	V	E	F	V	N	A	480					
QY	481	G	H	T	V	L	S	Q	V	H	F	A	B	E	D	T	G	I	S	L	Y	T	D	G	G	P	A	H	T	G	I	V	R	E	I	G	O	A	I	521																		
Db	481	G	H	T	V	L	S	Q	V	H	F	A	B	E	D	T	G	I	S	L	Y	T	D	G	G	P	A	H	T	G	I	V	R	E	I	G	O	A	I	521																		

RESULT 2

O50585

PRELIMINARY;

PRT;

517 AA.

ID

O50585;

AC

O50585;

DT

01-JUN-1998 (TrEMBLrel. 06, Created)

DT

01-JUN-1998 (TrEM

```
QY 294 GWDYAAVTPSIDAPETKRLAIAMNNMKYAAARDVPTDASDGYNGQNSIVRELRLARQP 353
Db 303 GWDYAAVTPKAEAPETRLAIAMNNMKYAAARDVPTDASDGYNGQNSITRELRLERQE 362
QY 354 GGWYTLISFPVAALTYVATTTLPTDRTVDGSAVLFPNGRAYEIELDIADWTATNVGIVS 413
Db 363 GGWYSLSPFVAGLGYVATTIPTQVYNGSVLVFPWSGRAYEIELDISWTDAAVGLSV 422
QY 414 GRSPDCTRTITNIGKYGADLYVDRGSPDLGAGYSLAPYSRAAAPIDPCGARSVHLRLIYLDQTS 473
Db 423 GRSADGARTHTNIGKYGADLYVDRGSPDLGAGYSLAPYSRAAAPIDPCGARSVHLRLIYLDQTS 482
QY 474 VEVFNAGHTVLSQQVHFAGEGDTGISLYTDGGPAHFTGIVVREIGQAI 521
Db 483 VEVFNAGHTVLSQQVHFAGEGDTGISLYTDGGPAHFTGIVVREIGQAI 530
RESULT 4
QY3R69 ID QY3R69 PRELIMINARY; PRT; 620 AA.
AC QY3R69;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Levanase.
OS Microbacterium laevaniformans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Micrococccineae; Microbacteriaceae; Microbacterium.
OX NCBI_TaxID=36807;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15953;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunz F., Ogawara N., Mosser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borrius R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinol S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porvillik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Taseuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunz F., Ogawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; Z94043; CAB08014.1;
DR EMBL; Z99121; CAB15451.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR Glycosidase; Hydrolase;
KW Glycosidase; Hydrolase; Hypothetical protein; Complete proteome.
```

```
QY 403 WDTATNVGIVSGRSPDCTRTITNIGKYGADLYVDRGSPDLGAGYSLAPYS---RAAAPIDPG 459
Db 435 RQADNVGQLRSADGSHRDLTRDYLALNRAQTG-----RDSKWKVERTELD-D 487
QY 460 ARSVHLRLIYLDQTSVEVFNAGHTVLSQQVHFAGEGDTGISLYTDGGPAHFTGIVVRE 516
Db 488 ADTVLRLIYLDQTSVEVFNAGHTVLSQQVHFAGEGDTGISLYTDGGPAHFTGIVVRE 544
RESULT 5
ID07003 ID 007003 PRELIMINARY; PRT; 516 AA.
AC 007003;
DT 01-JUL-1997 (TremBLrel. 04, Created)
DT 01-JUL-1997 (TremBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical protein yveb.
GN YVEB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Denizot F.C.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunz F., Ogawara N., Mosser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borrius R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinol S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porvillik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Taseuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunz F., Ogawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; Z94043; CAB08014.1;
DR EMBL; Z99121; CAB15451.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR Glycosidase; Hydrolase; Hypothetical protein; Complete proteome.
```

	Query Match	30.9%; Score 880.5; DB 16; Length 514;
	Best Local Similarity	40.7%; Pred. No. 4.3e-50;
	Matches	200; Conservative 76; Mismatches 179; Indels 37; Gaps
Qy	40	RAVYHMTPPSGLCDPQPVTTHCAYOLYLHS--DQNGPG-GWDHASTTDGVAFHHGT 97 : : : : : :
Db	45	REVHTFVPNKWKNDPQPVFFDEGYHYLYLNKDYPNGNGETWRQATSKDLVTWKEGV 104 : : : : : :
Qy	98	VMP--LRPDFVWSGAVVGTAAGFGAGAVVALAQTPTDGVRKYQEQLYVSTDDGET 155 : : : : : :
Db	105	CIPKTYENGDIWSSFFVDAQNATAGFKGAIVAIVTPQSASMDK-QEQFLWYSTDRGT 163 : : : : : :
Qy	156	FTALPD-PVIYNTGGRATTPAEIENAEWFDPDKIHWDTARGEWCVIGRLRYAAFYTS 214 : : : : : :
Db	164	FKPSNQPIPN-----PCTKD----FRDPKIWIDEFKNKKVMVALEGTGIGEYESY 211 : : : : : :
Qy	215	NLROWTLRRNFDPYNHALGGTECPDFEITADGTRHWLAASMDAYGICLPMTYAWTC 274 : : : : : :
Db	212	NLKWN--OHTGDFTNNIGIVECPDIEMFMSNDGNCKWIIJGTSANGKVSGEPNTYAYWG 269 : : : : : :
Qy	275	TWDGEOPHADDLTPOWLDMGWDAVAAVWPVSIDAPET--KRLAIAMNNKFYAARDVPTD 332 : : : : : :
Db	270	NYDGCKFTADISTPKWLDFGFDWAATAVFSENGSKNLKRYALAWNKHWDY-ANNTPT- 327 : : : : : :
Qy	333	ASDYGNGONSIVRELRLAROPGGWYTLLSPVAALTINYVTATTLPDRTVDGSVIPWNG 392 : : : : : :
Db	328	IQNRFMGDSIVREITLSKQKDNYSLLSKSPIKRNIETTSDIQFKQISVKGLKLHKVQG 387 : : : : : :
Qy	393	RAYELDIAMDYATNVGISVGRSPDGRTHNICKY--GADLVYDRGPSDLACYSLAP-- 448 : : : : : :
Db	388	KAYQLDTINWVDKAKNGVIRLESDDRHRIDGIFTEGKYSVNR-----AYTGPNPK 441 : : : : : :
Qy	449	--YSRAAAPIDPCARSVHLRILVDYOTSVEFVNAGHTVLSSQQVHFAGEGTGTSYTDGG 505 : : : : : :
Db	442	SKKYVESRAPDINNKKVHLRIEFVDKYSVEVFIDDGKITYSNEVFPPEDPDKGITLSIKG 501 : : : : : :
Qy	506	PAHFTGIIVREI 517
Db	502	KAVEKNITIKHI 513 : : : : : :

```

RESULT 6
Q97180 PRELIMINARY; PRT; 514 AA.
AC Q97180;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Levanase.
GN CAC1773;
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusev R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT *Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum*;
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007686; AAK79738.1;
DR InterPro: IPR001362; GH32.
DR Pfam: PF00251; Glyco_hydro_32; 1.
KW Complete proteome.
SQ SEQUENCE 514 AA; 59636 MW; 50366383B2D97004 CRC64;

```

Db	73	PPAIPPGEDGAIFSGSAVVDKNNITSGFDEEGSG-LVAYITNBNKNSQP-----GKPVQVS	127
Qy	146	LYWSTDGSGFTTALP-DPVIVNTDGRAATTPAETIENAEFRDPKIHWDTARGEWCVIGR	204
Db	128	IAYSKOKGRWTYEGNPVLFPTDILD-----FRDPKVIWHDSSKWIWLVAV	175
Qy	205	LRYAAYTSPNLRDWTLRRNF--DYPNHALGGIECPDLFEITADD--GTRHWVLAASM--	258
Db	176	RDRVEFTSPNLKWSFASFEFGSDIPHIHRGIFECPDIFRIQVDEDLNTTKWILMLSVGD	235
Qy	259	-----DAYGIGLPMYAYWGTWDEGFHADDL-----TPQWLWDGWDWYAAVT	302
Db	236	RNGYNPDPPPPAGSGSM---MYFVGSFGDKSFTPEALESIDTIKKWDYDGSDFYAAS	291
Qy	303	WPSIDAPETKRLATAWNKNKYAARDVPTDASDGYNGCONSIIVRELRLAROPGGWYTLT	362
Db	292	WNGISNEDGRRKILGWMNNWRYAT---TLPSKEWGRKTSIPRELQRLYRPEG-LRLQ	346
Qy	363	PVAALTNVYVATTTLPTDPTV-DGSAVL-PWNGRAYETIELDIAWDATNVGISYGRSPDGT	420
Db	347	PINLSQLRPILSLQDLTIKPGMNVLSDISAAKAEITAEFIGTAVEFGKVRKS--AN	404
Qy	421	RHTNIGK--YGADLYVDRGSDLAGYSLAPYSRAAAPIDPGARSVHLRLIVDTQSVEFV	478
Db	405	QETIIIGNISNEELFVDRTKSSATDFHSDFTAIHKATMKPEHERIQLSYLDWSSVEFG	464
Qy	479	NAGHTVLSQOVHFAEGDTGSLYTDGGPAHFTGIVVREI	517
Db	465	NHGKAIIISDIFDPFESKGLELVAIGGELRVVSLQINDL	503
RESULT 9			
Q9LAL1	ID	Q9LAL1	PRELIMINARY; PRT; 501 AA.
AC	Q9LAL1		
DT	01-OCT-2000	(TrEMBLrel. 15, Created)	
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)	
DT	01-MAR-2002	(TrEMBLrel. 20, Last annotation update)	
DE	Exo-inulinase.		
GN	INU2.		
OS	Pseudomonas mucidolens.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;		
OC	Pseudomonas.		
OX	NCBI_TaxID=46679;		
RN	[1].		
RP	SEQUENCE FROM N.A.		
RT	Kwon Y.-M., Choi Y.-J.:		
RT	"DNA sequences and expression in Escherichia coli of an exo-inulinase		
RL	gene (inu2) from Pseudomonas mucidolens";		
RL	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.		
CC	-/- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.		
DR	EMBL: AF129819; AAF44125.1; -		
DR	Interpro: IPR001362; GH_32.		
DR	Pfam: PF00251; Glyco_hydro_32.1.		
DR	PROSITE: PS00609; GLYCOSYL_HYDROL_F32; 1.		
KW	Glycosidase; Hydrolase.		
SQ	SEQUENCE 501 AA; 57797 MW; 4043B86F20B4A4E7 CRC64;		
Query Match 18.5%; Score 527.5; DB 2; Length 501;			
Best Local Similarity 31.3%; Pred. No. 7.4e-27;			
Matches 156; Conservative 72; Mismatches 223; Indels 47; Gaps			
Qy	40	RAVYHMTPPSGWLCDDPQPVYTHGAYQLYHLSDDNN--GPGGWDHASTTGDVAETHHGT	97
Db	11	RPOFHFTETNMNDPNGLVYEGEYHLFTQYHPFGNTGMPHMGHVAVKDLIHWDD--L	68
Qy	98	VMPLRPDF--PVMSGSAVVGTAANTAGF--GAGAVVALATQP---TDGVRKIQEQLYWS	150
Db	69	PIALYPHIQIFSGSAVDQWQTSGFNGAGMVALFTHADKFPDSGRPQRLAYST	128
Qy	151	DGGFTTALPDPVIVNTDGRAATTPAETIENAEFRDPKIHWDTARGEWCVIGRLRYAAF	210

```

Db 129 DRGRW-----IKYGNPFDAEAMTD---FRDPKVFVHKASKNKIMILAAGNCVRI 177
Qy 211 YTSPLRDLTLRRNPDYENHALGGI-ECPDLFETADGTRH---WVLAASW---DAYGI 263
Db 178 YHSSNLKQWKASERGANEGSHGTGWCEDLFLPV-DGDRNRKKVWVVSIGNSDEYLE 236
Qy 264 GLPMTYAYWTGTWGEQF--HADDLTQWLDGWDWYAAVTPSIDAPETKRLAIAMNN 321
Db 237 G--SRTQVFIGOFDGIREFTNENDPETVLWVDHGRDNYAGVTSWDIPEEDGRRLFIAMSN 294
Qy 322 WKYAADVPTDASDGYNGQNSIVRELRLARQPGGWYLLSTPVAALTYNYVATTLPDR- 380
Db 295 WKY-ANHTPTET---WRSAMTIPRSLSRSTPEG-TRLFQTPVTELESRLRLEWKDLE 349
Qy 381 -TVDSAVLPWNGRAYETELEDIAWDTATNVGISVGRSPDGRHTNIGYKAD---LYVDR 436
Db 350 VTTNGVNLNSIGDIFEIIAEFELNTALEFGFKVKSAD--QETIVG-YDAEQOMLFIDR 406
Qy 437 GPSDLAGYSLAPYRAAAPIPDGARSVHLRLVLTQSVVEFVNAAGHTVLSOOVHFAEGDT 496
Db 407 TQSGVSDCEHPCKHGAVMIPNQNRIOHMFVDRSSVEFGNGELTMTDLIPDDSSST 466
Qy 497 GISLYTDCGPAHFTGIVV 514
Db 467 GIEVTAKEGNVKLVSML 484

RESULT 10
ID Q45155 PRELIMINARY; PRT; 622 AA.
AC Q45155;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Levansae.
GN SCRL.
OS Bacteroides fragilis.
OC Bacteria; CFB group; Bacteroidetes; Bacteroidales; Bacteroidaceae;
OC Bacteroides.
OX NCBI_TaxID=817;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BF1;
RX MEDLINE=93259952; PubMed=8491724;
RA Blatch G.L., Woods D.R.;
RT "Molecular characterization of a fructanase produced by Bacteroides fragilis BF-1."
RL J. Bacteriol. 175:3058-3066(1993).
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; M83774; AAA2924.1;
DR InterPro; IPR001362; GH_32.
KW Pfam; PF00251; Glyco_hydro_32; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 622 AA; 70287 MW; 63424EB274CEE134 CRC64;

Query Match 18.18; Score 515; DB 2; Length 622;
Best Local Similarity 30.24; Pred. No. 6.6e-26;
Matches 156; Conservative 77; Mismatches 195; Indels 88; Gaps 20;

Qy 40 RAVYHMTPPSCWGLCDPQRPVTHTCAYOLYLHSDQNN--GPGGWDHASTTGDVAFTHGT 97
Db 133 RPLVHTHTPLYGWMDANGLVTKGDEYHLFYQYNPYGSMGNMHWGHSVKDLVHWEH--- 189
Qy 98 VWLPRDPFP-----VMSGSVAVGTATAGFAGAVVALATOPTDGVGRKYQEOQLYVSTD 151
Db 190 ---LEPALARDTLGHIFSGSVDDANTAGYAGAIYFTSAD--KNGIOCMAYSTD 244
Qy 152 GGFTFTAL-PDPVIVNTDGRAATTPABIEAENWFRDPKIHWDTARGEWCVIGRLRYAAF 210
Db 245 NGRFTFYKPNVLTPTD-----LKDRPQKVEWYAPDQKQWVWVVSADKEMRF 293
Qy 211 YTSPLRDLTLRRNFD-----YPNHALGGTECPDLFEITADGTRH---WVLAASMDAYGI 263

```

```

Db 294 YSENKLEWTVMSGWEGYGVQPSQ-----FECPDVVELPVDGNPDHKKWALIVNVN---- 345
Qy 264 GLPMTY-----AYWTGTWGEQFHADDL--TPQWLDGWDWYAAVTPSIDAPETKRLA 315
Db 346 --PCYFGGSATQYFIGDFGKFKVCDNKPETVWLDGKDHATVCFSTNG---DRTIA 400
Qy 316 IAMNNKYYAARDVPTDASDGYNGQNSIVRELRLARQPGGWYLLSTPVAALTYNYVATT 375
Db 401 VPMNSNQY-ANIVPTRQ---FRSANALPRELSLYTQGDYI-NAAPVEETKSLRKESR 455
Qy 376 TLPDRTVDGS-----AVLPWNGRAYETELEDIAWDTATNVGISVGRSPDGRHTNIGYKAD 431
Db 456 EIPAFEVGDYHVDLSLSDNKGAYEIELELAAGSAEIMGLKLFNEKGENVDYIYSLPEKK 515
Qy 432 LYVDRGPDSLA--CYSLAPYSRAA-----APIDPGARSVHLRLTV 469
Db 516 LVMDRTSKGVDFGKDSAPHAIEAHRKQNSINYVDFAELGTWAPVQK-AGNYKLDIFV 574
Qy 470 DTQSVVEFVNAAGHTVLSOOVHFAEGDTGISLYTDCG 505
Db 575 DKCSVEIFLNGKLTAMTNLIPPTTPYQMSFYSRGG 610

RESULT 11
Q9RBJ1
ID Q9RBJ1 PRELIMINARY; PRT; 534 AA.
AC Q9RBJ1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Levansae precursor.
GN LSDB.
OS Acetobacter diazotrophicus (Gluconacetobacter diazotrophicus).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.
OX NCBI_TaxID=33996;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SRT4;
RX MEDLINE=96253999; PubMed=8704949;
RA Arrieta J., Hernandez L., Coego A., Suarez V., Balmori P.,
Menendez C., Petit-Glatron M.F., Chambert R., Selman-Housein G.;
RT "Molecular characterization of the levansucrase gene from the
endophytic sugarcane bacterium Acetobacter diazotrophicus SRT4."
RL Microbiology 142:1077-1085(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SRT4;
RX MEDLINE=99373266; PubMed=10441728;
RA Hernandez L., Arrieta J., Betancourt L., Falcon V., Madrazo J.,
Coego A., Menendez C.;
RT "Levansucrase from Acetobacter diazotrophicus SRT4 is secreted via
periplasm by a signal-peptide-dependent pathway."
RL Curr. Microbiol. 39:146-152(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SRT4;
RA Menendez C., Hernandez L., Mendoza M.F., Hevia P., Selman G.,
Arrieta J.;
RT "Molecular cloning and expression in E. coli of an exolevanase from
Acetobacter diazotrophicus SRT4."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; L41732; AAF16405.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 36 POTENTIAL.
FT CHAIN 37 534 LEVANASE.
SQ SEQUENCE 534 AA; 58367 MW; 7BDBA4A9BA478F6 CRC64;

Query Match 18.08; Score 513; DB 2; Length 534;

```


Db 133 IYTSFAKESGVHQTQOASLAFSTDAAGMTWS-----KYAGNPVLGRGSAHFRD 180
Qy 187 PKI--HMDTARGEVWCVIGRLRY--AAFYTSPLNRLDWTLRNEDYPNHALGGIECPDLFE 242
Db 181 PKFVRYEGPAGSRWIMVAVEAQHOQVVLVYRSADLKDWYLSLTFGPANASDGEWECPP--LP 238
Qy 243 ITADGGRHWVLAASMDAYGIGLPMYVYWTGTWDSQFHA--DDLTPQ----- 289
Db 239 VDGPDNVKVLVNVINPGAVAGSGGQYFVGDFGVQFPADPSLPTDADGNVDLRHC 298
Qy 290 -WLDGWDWYAAYTWPISIDAPETKRLAIAWMNNKYYAARDVPTDASDGYNGQNSIVRELRL 348
Db 299 LWDGWRDYYAAVSFS--NAPENRRIMIGWNNWDY--ANSLPTSP--WRSSMSLAREIE 352
Qy 349 LARQPGGWYTLSTPVAALTYVTATT-----TLPDRTVDCGSVLPWNGRAYEI 397
Db 353 LA--TVDGFPRLVQRPVLPDSGEPARTIONMELTTPCCNCPTRRPPGSAQL-----I 403
Qy 398 ELDIAMDTATNVGISVGRSPDGTHT--NIGKYGADLYVDRGPSDLAGYSLAPYSRAAAP 455
Db 404 EAEILPCTARHIAFRLLGAPDGSAAATVLSFDALTSRLTLDRNSGNTAFTISSRRLNPQ 463
Qy 456 IDPG-ARSVHLRIILVDTQSVVEF 477
Db 464 VKASTGGALRLKVIIDQCSVEF 486

RESULT 14

Q93T55 ID Q93T55 PRELIMINARY; PRT; 493 AA.
AC Q93T55
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Exonulinase (EC 3.2.1.80).
GN EXOL.
OS Bacillus sp. snu-7.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=159251;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SNU-7;
RA Koo B.-S., Pan C.-H., Kim S.-I.;
RT "Molecular Cloning of Exonulinase Gene from Bacillus sp. snu-7 and
its Expression in Escherichia coli."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL: AF366292; AAK54126.1;
DR InterPro: IPR001362; GH_32.
DR Pfam: PF00251; Glyco_Hydro_32; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 493 AA; 53544 MW; EC99C9976BD0E46 CRC64;

Query Match 16.5%; Score 468.5; DB 2; Length 493;
Best Local Similarity 29.4%; Pred. No. 5.6e-23;
Matches 148; Conservative 70; Mismatches 212; Indels 73; Gaps 18;

Qy 15 GAGALALIFGAGVPPARASAPGSLRAVYHMTPPSGHLCDPQRPVTHGAYQLYLSHQ 74
Db 17 GYGLKRL--GTAASLAAITVG--RPLHYTAKNTLNDPGLVHQQGVYHLFYQNPF 72
Qy 75 NN--GPGGWDHASTDGVAFTHGTVPLRDPFVWSGSAVVGTVANTAGFCAGAVVALAT 132
Db 73 DNVWGNWSGWHATSTDLHLHTEHPVATACNEEDVFSGSIVDGHGTSFGTEADPALVA 132
Qy 133 OPTDGVK-----YQBYLYWSTDGGFTTALPDVIVNTDGRAATTPAEIENAEWFRD 186
Db 133 IYTSFAKESGVHQTQOASLAFSTDAAGMTWS-----KYAGNPVLGPGSAHFRD 180
Qy 187 PKI--HMDTARGEVWCVIGRLRY--AAFYTSPLNRLDWTLRNEDYPNHALGGIECPDLFE 242

Db 181 PKFVRYEGPAGSRWIMVAVEAQHOQVVLVYRSADLKDWYLSLTFGPANASDGEWECPP--LP 238
Qy 243 ITADGGRHWVLAASMDAYGIGLPMYVYWTGTWDSQFHA--DDLTPQ----- 289
Db 239 VDGPDNVKVLVNVINPGAVAGSGGQYFVGDFGVQFPADPSLPTDADGNVDLRHC 298
Qy 290 -WLDGWDWYAAYTWPISIDAPETKRLAIAWMNNKYYAARDVPTDASDGYNGQNSIVRELRL 348
Db 299 LWDGWRDYYAAVSFS--NAPENRRIMIGWNNWDY--ANSLPTSP--WRSSMSLAREIE 352
Qy 349 LARQPGGWYTLSTPVAALTYVTATT-----TLPDRTVDCGSVLPWNGRAYEI 397
Db 353 LA--TVDGFPRLVQRPVLPDSGEPARTIONMELTTPCCNCPTRRPPGSAQL-----I 403
Qy 398 ELDIAMDTATNVGISVGRSPDGTHT--NIGKYGADLYVDRGPSDLAGYSLAPYSRAAAP 455
Db 404 EAEILPCTARHIAFRLLGAPDGSAAATVLSFDALTSRLTLDRNSGNTAFTISSRRLNPQ 463
Qy 456 IDPG-ARSVHLRIILVDTQSVVEF 477
Db 464 VKASTGGALRLKVIIDQCSVEF 486

RESULT 15

Q44109 ID Q44109 PRELIMINARY; PRT; 943 AA.
AC Q44109
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Levansase.
GN LEVJ.
OS Actinomycetes naeslundii.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=1655;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T14V;
RX MEDLINE=95129923; PubMed=7828936;
RA Norman J.M., Bunney K.L., Giffard P.M.;
RT "Characterization of levJ, a sucrose/fructanase-encoding gene from
Actinomycetes naeslundii T14V, and comparison of its product with other
sucrose-cleaving enzymes.";
RL Gene 152:93-98(1995).
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL: U12274; AAA67876.1;
DR InterPro: IPR001362; GH_32.
DR Pfam: PF00251; Glyco_Hydro_32; 1.
DR PROSITE: PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 943 AA; 99267 MW; 801A2A7D074579B9 CRC64;

Query Match 16.4%; Score 467.5; DB 2; Length 943;
Best Local Similarity 21.5%; Pred. No. 1.6e-22;
Matches 187; Conservative 83; Mismatches 217; Indels 381; Gaps 27;

Qy 4 AISRRVLAQAGAGALALIFGG----AVPPAARASAPGS--LRVYHMTPPSGWLCDPQRP 58
Db 19 AASTLASMPAASGQARPVSAAAPAPAPKADQGTGERWRPQSHYTPKNNMNDPGL 78
Qy 59 VTHGAYQLYLSHQNN--GPGGWDHASTDGVAFTHGTVPLRDPFVWSGSAVVGTV 116
Db 79 VYDGEYHMFYQYNPEGSWGNMWSGHAWSKOLVHMQELGVAIPHTSQYGVFSGSAIDT 138
Qy 117 ANTAGFCA---GAVVALATQPTDGVKRYQBYLYWSTDGGFTTAL--PDVIVNTDGRA 171
Db 139 KNTSGLSGSPDNPAMVAVWTRADYGGN--OSQSLAYSTDKGGTNLNNNGDPVL----- 189
Qy 172 ATPAEIENAEWFRDPIKIHWDTARGEWVCVIGRL--RYAAFYTSPLNRLDWTLRNEDYPN 229
Db 190 -----DIGSNE-FRDPKVFWDQASGRWTVVVSHTAEHRVSFYSPDLIHWTEQSGEG 243


```
QY 230 HALGIECPDLEITADDGTRH--WVLA----- 255
Db      ||| | : | : |||
244 ITSAYWACPDFPEPLVDGSSQEVKVLVVTVADSAQYFVGSWDGTTFTPEIPHYSGEG 303
QY 256 -----ASMDAYG----- 262
Db 304 TTLADFENYAGWKADGAAGFGSPATGDLPGHQKAYVDFSGGDADTGTLTSDFTVSS 363
QY 263 ----- 262
Db 364 SYINLRTAGGKHPYNQATGDNNGGRLLAGFDGSWEGTWVEGSAFAATPPQCATPAQOPL 423
QY 263 -----IGLPMY-----AY----- 271
Db 424 VNHSSAGLLNTYLDAAATQGGSDAPTGTATSPFTTIDSAYLNLMLGGGNNRPRGGADGGS 483
QY 272 -----W-----TGTW----- 276
Db 484 RVSVELIVDGKVRSATGRNLEELNMQSDVSDLKGSQIIVTDTATGGMGHILLDEV 543
QY 277 ----- 276
Db 544 RASDKKASPIADNTSVNLVVDGKVVASATGNSGTLEWTSNMVAAKGRKARLVIEDRNG 603
QY 277 -----DGEQFHADLTQWLDGWNDWYAAVTWPSIDAPETKRLAIAMNN 320
Db 604 NAEDWGHLMVDQILQSDTKAFSGADWVPR-LDYGKDYAAVTWNV--PNGKRYQVGWMS 660
QY 321 NWKYAARDVPTDASDGYNGQNSIVRELRLARQPGGWYTLSTPVAALTNVTA-----T 374
Db 661 NWAY-VROLPFTT---WRTAMSTVREMGLTR-VNGKRLTAQOPVTALESRTGOELIRKD 715
QY 375 TTLP-DRTVGSAVLPWNGRAYEIELDIADWTATNVGISVGRSPDGRHTNIG--KYGAD 431
Db 716 TDIPVGETSLGKAA---QGTSLDISVDLSPSASSFAGLKV--LDNGEQYTLIGYDSQAKQ 770
QY 432 LYVDRGPSDLAGYSLAPYSRAAAPIDPGAR-SVHLRIIVDTQSVFVFNAGHTVLSQQVH 490
Db 771 LVVDRTHSGVTDSPKFFPARSTAPLSPDSKGQVHLRIIVDAHSVEVFAADGTPVITQVY 830
QY 491 FAEGDTGISLYTDGGPAHFTGIVVREIG 518
Db 831 PEQDATGVSLYAEAGTAHLGSLSLHLG 858
```

Search completed: May 6, 2003, 12:34:41
Job time : 61.1096 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)